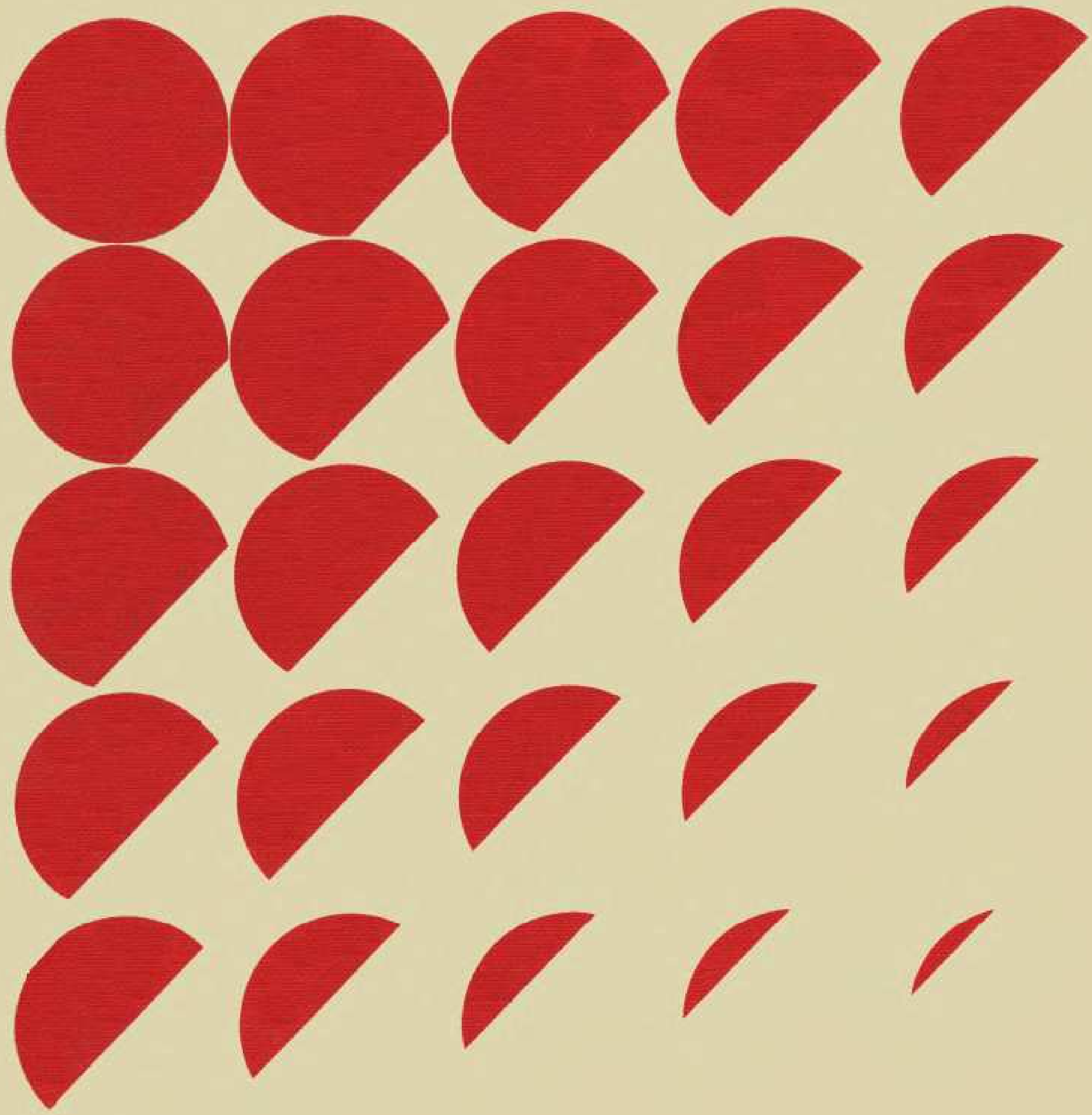
Hoel
Port
Stone

Introduction to Stochastic Processes



The Houghton Mifflin Series in Statistics
under the Editorship of Herman Chernoff

LEO BREIMAN
Probability and Stochastic Processes: With a View Toward Applications
Statistics: With a View Toward Applications

PAUL G. HOEL, SIDNEY C. PORT, AND CHARLES J. STONE
Introduction to Probability Theory
Introduction to Statistical Theory
Introduction to Stochastic Processes

PAUL F. LAZARSFELD AND NEIL W. HENRY
Latent Structure Analysis

GOTTFRIED E. NOETHER
Introduction to Statistics—A Fresh Approach

Y. S. CHOW, HERBERT ROBBINS, AND DAVID SIEGMUND
Great Expectations: The Theory of Optimal Stopping

I. RICHARD SAVAGE
Statistics: Uncertainty and Behavior

Introduction to Stochastic Processes

Paul G. Hoel
Sidney C. Port
Charles J. Stone
University of California, Los Angeles

HOUGHTON MIFFLIN COMPANY BOSTON
New York Atlanta Geneva, Illinois Dallas Palo Alto

PRINTED IN THE U.S.A.

LIBRARY OF CONGRESS CATALOG CARD NUMBER: 79-165035

ISBN: 0-395-12076-4

General Preface

This three-volume series grew out of a three-quarter course in probability, statistics, and stochastic processes taught for a number of years at UCLA. We felt a need for a series of books that would treat these subjects in a way that is well coordinated, but which would also give adequate emphasis to each subject as being interesting and useful on its own merits.

The first volume, *Introduction to Probability Theory*, presents the fundamental ideas of probability theory and also prepares the student both for courses in statistics and for further study in probability theory, including stochastic processes.

The second volume, *Introduction to Statistical Theory*, develops the basic theory of mathematical statistics in a systematic, unified manner. Together, the first two volumes contain the material that is often covered in a two-semester course in mathematical statistics.

The third volume, *Introduction to Stochastic Processes*, treats Markov chains, Poisson processes, birth and death processes, Gaussian processes, Brownian motion, and processes defined in terms of Brownian motion by means of elementary stochastic differential equations.

Preface

In recent years there has been an ever increasing interest in the study of systems which vary in time in a random manner. Mathematical models of such systems are known as stochastic processes. In this book we present an elementary account of some of the important topics in the theory of such processes. We have tried to select topics that are conceptually interesting and that have found fruitful application in various branches of science and technology.

A *stochastic process* can be defined quite generally as any collection of random variables $X(t)$, $t \in T$, defined on a common probability space, where T is a subset of $(-\infty, \infty)$ and is thought of as the time parameter set. The process is called a *continuous parameter process* if T is an interval having positive length and a *discrete parameter process* if T is a subset of the integers. If the random variables $X(t)$ all take on values from the fixed set $\mathscr{S}$, then $\mathscr{S}$ is called the *state space* of the process.

Many stochastic processes of theoretical and applied interest possess the property that, given the present state of the process, the past history does not affect conditional probabilities of events defined in terms of the future. Such processes are called *Markov processes*. In Chapters 1 and 2 we study *Markov chains*, which are discrete parameter Markov processes whose state space is finite or countably infinite. In Chapter 3 we study the corresponding continuous parameter processes, with the "Poisson process" as a special case.

In Chapters 4–6 we discuss continuous parameter processes whose state space is typically the real line. In Chapter 4 we introduce *Gaussian processes*, which are characterized by the property that every linear combination involving a finite number of the random variables $X(t)$, $t \in T$, is normally distributed. As an important special case, we discuss the *Wiener process*, which arises as a mathematical model for the physical phenomenon known as "Brownian motion."

In Chapter 5 we discuss integration and differentiation of stochastic processes. There we also use the Wiener process to give a mathematical model for "white noise."

In Chapter 6 we discuss solutions to nonhomogeneous ordinary differential equations having constant coefficients whose right-hand side is either a stochastic process or white noise. We also discuss estimation problems involving stochastic processes, and briefly consider the "spectral distribution" of a process.

This text has been designed for a one-semester course in stochastic processes. Written in close conjunction with *Introduction to Probability Theory*, the first volume of our three-volume series, it assumes that the student is acquainted with the material covered in a one-semester course in probability for which elementary calculus is a prerequisite.

Some of the proofs in Chapters 1 and 2 are somewhat more difficult than the rest of the text, and they appear in appendices to these chapters. These proofs and the starred material in Section 2.6 probably should be omitted or discussed only briefly in an elementary course.

An instructor using this text in a one-quarter course will probably not have time to cover the entire text. He may wish to cover the first three chapters thoroughly and the remainder as time permits, perhaps discussing those topics in the last three chapters that involve the Wiener process. Another option, however, is to emphasize continuous parameter processes by omitting or skimming Chapters 1 and 2 and concentrating on Chapters 3–6. (For example, the instructor could skip Sections 1.6.1, 1.6.2, 1.9, 2.2.2, 2.5.1, 2.6.1, and 2.8.) With some aid from the instructor, the student should be able to read Chapter 3 without having studied the first two chapters thoroughly. Chapters 4–6 are independent of the first two chapters and depend on Chapter 3 only in minor ways, mainly in that the Poisson process introduced in Chapter 3 is used in examples in the later chapters. The properties of the Poisson process that are needed later are summarized in Chapter 4 and can be regarded as axioms for the Poisson process.

The authors wish to thank the UCLA students who tolerated preliminary versions of this text and whose comments resulted in numerous improvements. Mr. Luis Gorostiza obtained the answers to the exercises and also made many suggestions that resulted in significant improvements. Finally, we wish to thank Mrs. Ruth Goldstein for her excellent typing.

Table of Contents

1 Markov Chains

Consider a system that can be in any one of a finite or countably infinite number of states. Let $\mathscr{S}$ denote this set of states. We can assume that $\mathscr{S}$ is a subset of the integers. The set $\mathscr{S}$ is called the *state space* of the system. Let the system be observed at the discrete moments of time $n = 0, 1, 2, \ldots$, and let X_n denote the state of the system at time n.

Since we are interested in non-deterministic systems, we think of X_n, $n \geq 0$, as random variables defined on a common probability space. Little can be said about such random variables unless some additional structure is imposed upon them.

The simplest possible structure is that of independent random variables. This would be a good model for such systems as repeated experiments in which future states of the system are independent of past and present states. In most systems that arise in practice, however, past and present states of the system influence the future states even if they do not uniquely determine them.

Many systems have the property that given the present state, the past states have no influence on the future. This property is called the *Markov property*, and systems having this property are called *Markov chains*. The Markov property is defined precisely by the requirement that

$$(1) \quad P(X_{n+1} = x_{n+1} \mid X_0 = x_0, \ldots, X_n = x_n) = P(X_{n+1} = x_{n+1} \mid X_n = x_n)$$

for every choice of the nonnegative integer n and the numbers $x_0, \ldots, x_{n+1}$, each in $\mathscr{S}$. The conditional probabilities $P(X_{n+1} = y \mid X_n = x)$ are called the *transition probabilities* of the chain. In this book we will study Markov chains having *stationary* transition probabilities, i.e., those such that $P(X_{n+1} = y \mid X_n = x)$ is independent of n. From now on, when we say that X_n, $n \geq 0$, forms a Markov chain, we mean that these random variables satisfy the Markov property and have stationary transition probabilities.

The study of such Markov chains is worthwhile from two viewpoints. First, they have a rich theory, much of which can be presented at an elementary level. Secondly, there are a large number of systems arising in practice that can be modeled by Markov chains, so the subject has many useful applications.

In order to help motivate the general results that will be discussed later, we begin by considering Markov chains having only two states.

1.1. Markov chains having two states

For an example of a Markov chain having two states, consider a machine that at the start of any particular day is either broken down or in operating condition. Assume that if the machine is broken down at the start of the nth day, the probability is p that it will be successfully repaired and in operating condition at the start of the $(n + 1)$th day. Assume also that if the machine is in operating condition at the start of the nth day, the probability is q that it will have a failure causing it to be broken down at the start of the $(n + 1)$th day. Finally, let $\pi_0(0)$ denote the probability that the machine is broken down initially, i.e., at the start of the 0th day.

Let the state 0 correspond to the machine being broken down and let the state 1 correspond to the machine being in operating condition. Let X_n be the random variable denoting the state of the machine at time n. According to the above description

$$P(X_{n+1} = 1 \mid X_n = 0) = p,$$

$$P(X_{n+1} = 0 \mid X_n = 1) = q,$$

and

$$P(X_0 = 0) = \pi_0(0).$$

Since there are only two states, 0 and 1, it follows immediately that

$$P(X_{n+1} = 0 \mid X_n = 0) = 1 - p,$$

$$P(X_{n+1} = 1 \mid X_n = 1) = 1 - q,$$

and that the probability $\pi_0(1)$ of being initially in state 1 is given by

$$\pi_0(1) = P(X_0 = 1) = 1 - \pi_0(0).$$

From this information, we can easily compute $P(X_n = 0)$ and $P(X_n = 1)$. We observe that

$$\begin{aligned}
P(X_{n+1} = 0) &= P(X_n = 0 \text{ and } X_{n+1} = 0) + P(X_n = 1 \text{ and } X_{n+1} = 0) \\
&= P(X_n = 0)P(X_{n+1} = 0 \mid X_n = 0) \\
&\quad + P(X_n = 1)P(X_{n+1} = 0 \mid X_n = 1) \\
&= (1 - p)P(X_n = 0) + qP(X_n = 1) \\
&= (1 - p)P(X_n = 0) + q(1 - P(X_n = 0)) \\
&= (1 - p - q)P(X_n = 0) + q.
\end{aligned}$$

Now $P(X_0 = 0) = \pi_0(0)$, so

$$P(X_1 = 0) = (1 - p - q)\pi_0(0) + q$$

and

$$\begin{aligned} P(X_2 = 0) &= (1 - p - q)P(X_1 = 0) + q \\ &= (1 - p - q)^2\pi_0(0) + q[1 + (1 - p - q)]. \end{aligned}$$

It is easily seen by repeating this procedure n times that

$$(2) \qquad P(X_n = 0) = (1 - p - q)^n\pi_0(0) + q \sum_{j=0}^{n-1} (1 - p - q)^j.$$

In the trivial case $p = q = 0$, it is clear that for all n

$$P(X_n = 0) = \pi_0(0) \qquad \text{and} \qquad P(X_n = 1) = \pi_0(1).$$

Suppose now that $p + q > 0$. Then by the formula for the sum of a finite geometric progression,

$$\sum_{j=0}^{n-1} (1 - p - q)^j = \frac{1 - (1 - p - q)^n}{p + q}.$$

We conclude from (2) that

$$(3) \qquad P(X_n = 0) = \frac{q}{p + q} + (1 - p - q)^n \left(\pi_0(0) - \frac{q}{p + q}\right),$$

and consequently that

$$(4) \qquad P(X_n = 1) = \frac{p}{p + q} + (1 - p - q)^n \left(\pi_0(1) - \frac{p}{p + q}\right).$$

Suppose that p and q are neither both equal to zero nor both equal to 1. Then $0 < p + q < 2$, which implies that $|1 - p - q| < 1$. In this case we can let $n \to \infty$ in (3) and (4) and conclude that

$$(5) \qquad \lim_{n\to\infty} P(X_n = 0) = \frac{q}{p + q} \qquad \text{and} \qquad \lim_{n\to\infty} P(X_n = 1) = \frac{p}{p + q}.$$

We can also obtain the probabilities $q/(p + q)$ and $p/(p + q)$ by a different approach. Suppose we want to choose $\pi_0(0)$ and $\pi_0(1)$ so that $P(X_n = 0)$ and $P(X_n = 1)$ are independent of n. It is clear from (3) and (4) that to do this we should choose

$$\pi_0(0) = \frac{q}{p + q} \qquad \text{and} \qquad \pi_0(1) = \frac{p}{p + q}.$$

Thus we see that if X_n, $n \geq 0$, starts out with the initial distribution

$$P(X_0 = 0) = \frac{q}{p + q} \qquad \text{and} \qquad P(X_0 = 1) = \frac{p}{p + q},$$

then for all n

$$P(X_n = 0) = \frac{q}{p+q} \quad \text{and} \quad P(X_n = 1) = \frac{p}{p+q}.$$

The description of the machine is vague because it does not really say whether X_n, $n \geq 0$, can be assumed to satisfy the Markov property. Let us suppose, however, that the Markov property does hold. We can use this added information to compute the joint distribution of $X_0, X_1, \ldots, X_n$.

For example, let $n = 2$ and let x_0, x_1, and x_2 each equal 0 or 1. Then

$$\begin{aligned}
&P(X_0 = x_0, X_1 = x_1, \text{ and } X_2 = x_2) \\
&\quad = P(X_0 = x_0 \text{ and } X_1 = x_1)P(X_2 = x_2 \mid X_0 = x_0 \text{ and } X_1 = x_1) \\
&\quad = P(X_0 = x_0)P(X_1 = x_1 \mid X_0 = x_0)P(X_2 = x_2 \mid X_0 = x_0 \text{ and } X_1 = x_1).
\end{aligned}$$

Now $P(X_0 = x_0)$ and $P(X_1 = x_1 \mid X_0 = x_0)$ are determined by p, q, and $\pi_0(0)$; but without the Markov property, we cannot evaluate $P(X_2 = x_2 \mid X_0 = x_0 \text{ and } X_1 = x_1)$ in terms of p, q, and $\pi_0(0)$. If the Markov property is satisfied, however, then

$$P(X_2 = x_2 \mid X_0 = x_0 \text{ and } X_1 = x_1) = P(X_2 = x_2 \mid X_1 = x_1),$$

which is determined by p and q. In this case

$$\begin{aligned}
&P(X_0 = x_0, X_1 = x_1, \text{ and } X_2 = x_2) \\
&\qquad = P(X_0 = x_0)P(X_1 = x_1 \mid X_0 = x_0)P(X_2 = x_2 \mid X_1 = x_1).
\end{aligned}$$

For example,

$$\begin{aligned}
&P(X_0 = 0, X_1 = 1, \text{ and } X_2 = 0) \\
&\qquad = P(X_0 = 0)P(X_1 = 1 \mid X_0 = 0)P(X_2 = 0 \mid X_1 = 1) \\
&\qquad = \pi_0(0)pq.
\end{aligned}$$

The reader should check the remaining entries in the following table, which gives the joint distribution of X_0, X_1, and X_2.

x_0	x_1	x_2	$P(X_0 = x_0, X_1 = x_1, \text{ and } X_2 = x_2)$
0	0	0	$\pi_0(0)(1 - p)^2$
0	0	1	$\pi_0(0)(1 - p)p$
0	1	0	$\pi_0(0)pq$
0	1	1	$\pi_0(0)p(1 - q)$
1	0	0	$(1 - \pi_0(0))q(1 - p)$
1	0	1	$(1 - \pi_0(0))qp$
1	1	0	$(1 - \pi_0(0))(1 - q)q$
1	1	1	$(1 - \pi_0(0))(1 - q)^2$

1.2. Transition function and initial distribution

Let X_n, $n \geq 0$, be a Markov chain having state space $\mathscr{S}$. (The restriction to two states is now dropped.) The function $P(x, y)$, $x \in \mathscr{S}$ and $y \in \mathscr{S}$, defined by

$$P(x, y) = P(X_1 = y \mid X_0 = x), \qquad x, y \in \mathscr{S}, \tag{6}$$

is called the *transition function* of the chain. It is such that

$$P(x, y) \geq 0, \qquad x, y \in \mathscr{S}, \tag{7}$$

and

$$\sum_y P(x, y) = 1, \qquad x \in \mathscr{S}. \tag{8}$$

Since the Markov chain has stationary probabilities, we see that

$$P(X_{n+1} = y \mid X_n = x) = P(x, y), \qquad n \geq 1. \tag{9}$$

It now follows from the Markov property that

$$P(X_{n+1} = y \mid X_0 = x_0, \ldots, X_{n-1} = x_{n-1}, X_n = x) = P(x, y). \tag{10}$$

In other words, if the Markov chain is in state x at time n, then no matter how it got to x, it has probability $P(x, y)$ of being in state y at the next step. For this reason the numbers $P(x, y)$ are called the *one-step transition probabilities* of the Markov chain.

The function $\pi_0(x)$, $x \in \mathscr{S}$, defined by

$$\pi_0(x) = P(X_0 = x), \qquad x \in \mathscr{S}, \tag{11}$$

is called the *initial distribution* of the chain. It is such that

$$\pi_0(x) \geq 0, \qquad x \in \mathscr{S}, \tag{12}$$

and

$$\sum_x \pi_0(x) = 1. \tag{13}$$

The joint distribution of $X_0, \ldots, X_n$ can easily be expressed in terms of the transition function and the initial distribution. For example,

$$\begin{aligned} P(X_0 = x_0, X_1 = x_1) &= P(X_0 = x_0)P(X_1 = x_1 \mid X_0 = x_0) \\ &= \pi_0(x_0)P(x_0, x_1). \end{aligned}$$

Also,

$$\begin{aligned} &P(X_0 = x_0, X_1 = x_1, X_2 = x_2) \\ &\qquad = P(X_0 = x_0, X_1 = x_1)P(X_2 = x_2 \mid X_0 = x_0, X_1 = x_1) \\ &\qquad = \pi_0(x_0)P(x_0, x_1)P(X_2 = x_2 \mid X_0 = x_0, X_1 = x_1). \end{aligned}$$

Since X_n, $n \geq 0$, satisfies the Markov property and has stationary transition probabilities, we see that

$$\begin{aligned} P(X_2 = x_2 \mid X_0 = x_0, X_1 = x_1) &= P(X_2 = x_2 \mid X_1 = x_1) \\ &= P(X_1 = x_2 \mid X_0 = x_1) \\ &= P(x_1, x_2). \end{aligned}$$

Thus

$$P(X_0 = x_0, X_1 = x_1, X_2 = x_2) = \pi_0(x_0)P(x_0, x_1)P(x_1, x_2).$$

By induction it is easily seen that

$$(14) \qquad P(X_0 = x_0, \ldots, X_n = x_n) = \pi_0(x_0)P(x_0, x_1) \cdots P(x_{n-1}, x_n).$$

It is usually more convenient, however, to reverse the order of our definitions. We say that $P(x, y)$, $x \in \mathscr{S}$ and $y \in \mathscr{S}$, is a *transition function* if it satisfies (7) and (8), and we say that $\pi_0(x)$, $x \in \mathscr{S}$, is an *initial distribution* if it satisfies (12) and (13). It can be shown that given any transition function P and any initial distribution π_0, there is a probability space and random variables X_n, $n \geq 0$, defined on that space satisfying (14). It is not difficult to show that these random variables form a Markov chain having transition function P and initial distribution π_0.

The reader may be bothered by the possibility that some of the conditional probabilities we have discussed may not be well defined. For example, the left side of (1) is not well defined if

$$P(X_0 = x_0, \ldots, X_n = x_n) = 0.$$

This difficulty is easily resolved. Equations (7), (8), (12), and (13) defining the transition functions and the initial distributions are well defined, and Equation (14) describing the joint distribution of $X_0, \ldots, X_n$ is well defined. It is not hard to show that if (14) holds, then (1), (6), (9), and (10) hold whenever the conditional probabilities in the respective equations are well defined. The same qualification holds for other equations involving conditional probabilities that will be obtained later.

It will soon be apparent that the transition function of a Markov chain plays a much greater rôle in describing its properties than does the initial distribution. For this reason it is customary to study simultaneously all Markov chains having a given transition function. In fact we adhere to the usual convention that by "a Markov chain having transition function P," we really mean the family of all Markov chains having that transition function.

1.3. Examples

In this section we will briefly describe several interesting examples of Markov chains. These examples will be further developed in the sequel.

Example 1. Random walk. Let $\xi_1, \xi_2, \ldots$ be independent integer-valued random variables having common density f. Let X_0 be an integer-valued random variable that is independent of the ξ_i's and set $X_n = X_0 + \xi_1 + \cdots + \xi_n$. The sequence X_n, $n \geq 0$, is called a *random walk*. It is a Markov chain whose state space is the integers and whose transition function is given by

$$P(x, y) = f(y - x).$$

To verify this, let π_0 denote the distribution of X_0. Then

$$\begin{aligned} P(X_0 &= x_0, \ldots, X_n = x_n) \\ &= P(X_0 = x_0, \xi_1 = x_1 - x_0, \ldots, \xi_n = x_n - x_{n-1}) \\ &= P(X_0 = x_0)P(\xi_1 = x_1 - x_0) \cdots P(\xi_n = x_n - x_{n-1}) \\ &= \pi_0(x_0)f(x_1 - x_0) \cdots f(x_n - x_{n-1}) \\ &= \pi_0(x_0)P(x_0, x_1) \cdots P(x_{n-1}, x_n), \end{aligned}$$

and thus (14) holds.

Suppose a "particle" moves along the integers according to this Markov chain. Whenever the particle is in x, regardless of how it got there, it jumps to state y with probability $f(y - x)$.

As a special case, consider a *simple random walk* in which $f(1) = p$, $f(-1) = q$, and $f(0) = r$, where p, q, and r are nonnegative and sum to one. The transition function is given by

$$P(x, y) = \begin{cases} p, & y = x + 1, \\ q, & y = x - 1, \\ r, & y = x, \\ 0, & \text{elsewhere.} \end{cases}$$

Let a particle undergo such a random walk. If the particle is in state x at a given observation, then by the next observation it will have jumped to state $x + 1$ with probability p and to state $x - 1$ with probability q; with probability r it will still be in state x.

Example 2. Ehrenfest chain. The following is a simple model of the exchange of heat or of gas molecules between two isolated bodies. Suppose we have two boxes, labeled 1 and 2, and d balls labeled $1, 2, \ldots, d$. Initially some of these balls are in box 1 and the remainder are in box 2. An integer is selected at random from $1, 2, \ldots, d$, and the ball labeled by that integer is removed from its box and placed in the opposite box. This procedure is repeated indefinitely with the selections being independent from trial to trial. Let X_n denote the number of balls in box 1 after the nth trial. Then X_n, $n \geq 0$, is a Markov chain on $\mathscr{S} = \{0, 1, 2, \ldots, d\}$.

The transition function of this Markov chain is easily computed. Suppose that there are x balls in box 1 at time n. Then with probability x/d the ball drawn on the $(n + 1)$th trial will be from box 1 and will be transferred to box 2. In this case there will be $x - 1$ balls in box 1 at time $n + 1$. Similarly, with probability $(d - x)/d$ the ball drawn on the $(n + 1)$th trial will be from box 2 and will be transferred to box 1, resulting in $x + 1$ balls in box 1 at time $n + 1$. Thus the transition function of this Markov chain is given by

$$P(x, y) = \begin{cases} \dfrac{x}{d}, & y = x - 1, \\ 1 - \dfrac{x}{d}, & y = x + 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Note that the Ehrenfest chain can in one transition only go from state x to $x - 1$ or $x + 1$ with positive probability.

A state a of a Markov chain is called an *absorbing state* if $P(a, a) = 1$ or, equivalently, if $P(a, y) = 0$ for $y \neq a$. The next example uses this definition.

Example 3. Gambler's ruin chain. Suppose a gambler starts out with a certain initial capital in dollars and makes a series of one dollar bets against the house. Assume that he has respective probabilities p and $q = 1 - p$ of winning and losing each bet, and that if his capital ever reaches zero, he is ruined and his capital remains zero thereafter. Let X_n, $n \geq 0$, denote the gambler's capital at time n. This is a Markov chain in which 0 is an absorbing state, and for $x \geq 1$

$$P(x, y) = \begin{cases} q, & y = x - 1, \\ p, & y = x + 1, \\ 0, & \text{elsewhere.} \end{cases} \tag{15}$$

Such a chain is called a *gambler's ruin chain* on $\mathscr{S} = \{0, 1, 2, \ldots\}$. We can modify this model by supposing that if the capital of the gambler increases to d dollars he quits playing. In this case 0 and d are both absorbing states, and (15) holds for $x = 1, \ldots, d - 1$.

For an alternative interpretation of the latter chain, we can assume that two gamblers are making a series of one dollar bets against each other and that between them they have a total capital of d dollars. Suppose the first gambler has probability p of winning any given bet, and the second gambler has probability $q = 1 - p$ of winning. The two gamblers play until one of them goes broke. Let X_n denote the capital of the first gambler at time n. Then X_n, $n \geq 0$, is a gambler's ruin chain on $\{0, 1, \ldots, d\}$.

Example 4. Birth and death chain. Consider a Markov chain either on $\mathcal{S} = \{0, 1, 2, \ldots\}$ or on $\mathcal{S} = \{0, 1, \ldots, d\}$ such that starting from x the chain will be at $x - 1$, x, or $x + 1$ after one step. The transition function of such a chain is given by

$$P(x, y) = \begin{cases} q_x, & y = x - 1, \\ r_x, & y = x, \\ p_x, & y = x + 1, \\ 0, & \text{elsewhere}, \end{cases}$$

where p_x, q_x, and r_x are nonnegative numbers such that $p_x + q_x + r_x = 1$. The Ehrenfest chain and the two versions of the gambler's ruin chain are examples of *birth and death chains.* The phrase "birth and death" stems from applications in which the state of the chain is the population of some living system. In these applications a transition from state x to state $x + 1$ corresponds to a "birth," while a transition from state x to state $x - 1$ corresponds to a "death."

In Chapter 3 we will study *birth and death processes.* These processes are similar to birth and death chains, except that jumps are allowed to occur at arbitrary times instead of just at integer times. In most applications, the models discussed in Chapter 3 are more realistic than those obtainable by using birth and death chains.

Example 5. Queuing chain. Consider a service facility such as a checkout counter at a supermarket. People arrive at the facility at various times and are eventually served. Those customers that have arrived at the facility but have not yet been served form a waiting line or queue. There are a variety of models to describe such systems. We will consider here only one very simple and somewhat artificial model; others will be discussed in Chapter 3.

Let time be measured in convenient periods, say in minutes. Suppose that if there are any customers waiting for service at the beginning of any given period, exactly one customer will be served during that period, and that if there are no customers waiting for service at the beginning of a period, none will be served during that period. Let ξ_n denote the number of new customers arriving during the nth period. We assume that $\xi_1, \xi_2, \ldots$ are independent nonnegative integer-valued random variables having common density f.

Let X_0 denote the number of customers present initially, and for $n \geq 1$, let X_n denote the number of customers present at the end of the nth period. If $X_n = 0$, then $X_{n+1} = \xi_{n+1}$; and if $X_n \geq 1$, then $X_{n+1} = X_n + \xi_{n+1} - 1$. It follows without difficulty from the assumptions on ξ_n, $n \geq 1$, that X_n, $n \geq 0$, is a Markov chain whose state space is the nonnegative integers and whose transition function P is given by

$$P(0, y) = f(y)$$

and

$$P(x, y) = f(y - x + 1), \qquad x \geq 1.$$

Example 6. Branching chain. Consider particles such as neutrons or bacteria that can generate new particles of the same type. The initial set of objects is referred to as belonging to the 0th generation. Particles generated from the nth generation are said to belong to the $(n + 1)$th generation. Let X_n, $n \geq 0$, denote the number of particles in the nth generation.

Nothing in this description requires that the various particles in a generation give rise to new particles simultaneously. Indeed at a given time, particles from several generations may coexist.

A typical situation is illustrated in Figure 1: one initial particle gives rise to two particles. Thus $X_0 = 1$ and $X_1 = 2$. One of the particles in the first generation gives rise to three particles and the other gives rise to one particle, so that $X_2 = 4$. We see from Figure 1 that $X_3 = 2$. Since neither of the particles in the third generation gives rise to new particles, we conclude that $X_4 = 0$ and consequently that $X_n = 0$ for all $n \geq 4$. In other words, the progeny of the initial particle in the zeroth generation become extinct after three generations.

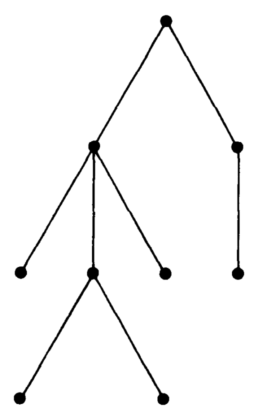


Figure 1

In order to model this system as a Markov chain, we suppose that each particle gives rise to ξ particles in the next generation, where ξ is a nonnegative integer-valued random variable having density f. We suppose that the number of offspring of the various particles in the various generations are chosen independently according to the density f.

Under these assumptions X_n, $n \geq 0$, forms a Markov chain whose state space is the nonnegative integers. State 0 is an absorbing state. For if there are no particles in a given generation, there will not be any particles in the next generation either. For $x \geq 1$

$$P(x, y) = P(\xi_1 + \cdots + \xi_x = y),$$

where $\xi_1, \ldots, \xi_x$ are independent random variables having common density f. In particular, $P(1, y) = f(y)$, $y \geq 0$.

If a particle gives rise to $\xi = 0$ particles, the interpretation is that the particle dies or disappears. Suppose a particle gives rise to ξ particles, which in turn give rise to other particles; but after some number of generations, all descendants of the initial particle have died or disappeared (see Figure 1). We describe such an event by saying that the descendants of the original particle eventually become *extinct*. An interesting problem involving branching chains is to compute the probability ρ of eventual extinction for a branching chain starting with a single particle or, equivalently, the probability that a branching chain starting at state 1 will eventually be absorbed at state 0. Once we determine ρ, we can easily find the probability that in a branching chain starting with x particles the descendants of each of the original particles eventually become extinct. Indeed, since the particles are assumed to act independently in giving rise to new particles, the desired probability is just ρ^x.

The branching chain was used originally to determine the probability that the male line of a given person would eventually become extinct. For this purpose only male children would be included in the various generations.

Example 7. Consider a gene composed of d subunits, where d is some positive integer and each subunit is either normal or mutant in form. Consider a cell with a gene composed of m mutant subunits and $d - m$ normal subunits. Before the cell divides into two daughter cells, the gene duplicates. The corresponding gene of one of the daughter cells is composed of d units chosen at random from the $2m$ mutant subunits and the $2(d - m)$ normal subunits. Suppose we follow a fixed line of descent from a given gene. Let X_0 be the number of mutant subunits initially present, and let X_n, $n \geq 1$, be the number present in the nth descendant gene. Then X_n, $n \geq 0$, is a Markov chain on $\mathscr{S} = \{0, 1, 2, \ldots, d\}$ and

$$P(x, y) = \frac{\binom{2x}{y}\binom{2d - 2x}{d - y}}{\binom{2d}{d}}.$$

States 0 and d are absorbing states for this chain.

1.4. Computations with transition functions

Let X_n, $n \geq 0$, be a Markov chain on $\mathscr{S}$ having transition function P. In this section we will show how various conditional probabilities can be expressed in terms of P. We will also define the n-step transition function of the Markov chain.

We begin with the formula

$$\begin{aligned}(16) \qquad P(X_{n+1} = x_{n+1}, \ldots, X_{n+m} = x_{n+m} \mid X_0 = x_0, \ldots, X_n = x_n)\\ = P(x_n, x_{n+1}) \cdots P(x_{n+m-1}, x_{n+m}).\end{aligned}$$

To prove (16) we write the left side of this equation as

$$\frac{P(X_0 = x_0, \ldots, X_{n+m} = x_{n+m})}{P(X_0 = x_0, \ldots, X_n = x_n)}.$$

By (14) this ratio equals

$$\frac{\pi_0(x_0)P(x_0, x_1) \cdots P(x_{n+m-1}, x_{n+m})}{\pi_0(x_0)P(x_0, x_1) \cdots P(x_{n-1}, x_n)},$$

which reduces to the right side of (16).

It is convenient to rewrite (16) as

$$\begin{aligned}(17) \quad P(X_{n+1} = y_1, \ldots, X_{n+m} = y_m \mid X_0 = x_0, \ldots, X_{n-1} = x_{n-1}, X_n = x)\\ = P(x, y_1)P(y_1, y_2) \cdots P(y_{m-1}, y_m).\end{aligned}$$

Let $A_0, \ldots, A_{n-1}$ be subsets of $\mathscr{S}$. It follows from (17) and Exercise 4(a) that

$$\begin{aligned}(18) \quad P(X_{n+1} = y_1, \ldots, X_{n+m} = y_m \mid X_0 \in A_0, \ldots, X_{n-1} \in A_{n-1}, X_n = x)\\ = P(x, y_1)P(y_1, y_2) \cdots P(y_{m-1}, y_m).\end{aligned}$$

Let $B_1, \ldots, B_m$ be subsets of $\mathscr{S}$. It follows from (18) and Exercise 4(b) that

$$\begin{aligned}(19) \quad P(X_{n+1} \in B_1, \ldots, X_{n+m} \in B_m \mid X_0 \in A_0, \ldots, X_{n-1} \in A_{n-1}, X_n = x)\\ = \sum_{y_1 \in B_1} \cdots \sum_{y_m \in B_m} P(x, y_1)P(y_1, y_2) \cdots P(y_{m-1}, y_m).\end{aligned}$$

The *m-step transition function* $P^m(x, y)$, which gives the probability of going from x to y in m steps, is defined by

$$(20) \qquad P^m(x, y) = \sum_{y_1} \cdots \sum_{y_{m-1}} P(x, y_1)P(y_1, y_2) \cdots P(y_{m-2}, y_{m-1})P(y_{m-1}, y)$$

for $m \geq 2$, by $P^1(x, y) = P(x, y)$, and by

$$P^0(x, y) = \begin{cases} 1, & x = y, \\ 0, & \text{elsewhere.} \end{cases}$$

We see by setting $B_1 = \cdots = B_{m-1} = \mathscr{S}$ and $B_m = \{y\}$ in (19) that

$$(21) \qquad P(X_{n+m} = y \mid X_0 \in A_0, \ldots, X_{n-1} \in A_{n-1}, X_n = x) = P^m(x, y).$$

In particular, by setting $A_0 = \cdots = A_{n-1} = \mathscr{S}$, we see that

$$(22) \qquad P(X_{n+m} = y \mid X_n = x) = P^m(x, y).$$

It also follows from (21) that

$$(23) \qquad P(X_{n+m} = y \mid X_0 = x, X_n = z) = P^m(z, y).$$

Since (see Exercise 4(c))

$$\begin{aligned} P^{n+m}(x, y) &= P(X_{n+m} = y \mid X_0 = x) \\ &= \sum_z P(X_n = z \mid X_0 = x)P(X_{n+m} = y \mid X_0 = x, X_n = z) \\ &= \sum_z P^n(x, z)P(X_{n+m} = y \mid X_0 = x, X_n = z), \end{aligned}$$

we conclude from (23) that

$$(24) \qquad P^{n+m}(x, y) = \sum_z P^n(x, z)P^m(z, y).$$

For Markov chains having a finite number of states, (24) allows us to think of P^n as the nth power of the matrix P, an idea we will pursue in Section 1.4.2.

Let π_0 be an initial distribution for the Markov chain. Since

$$\begin{aligned} P(X_n = y) &= \sum_x P(X_0 = x, X_n = y) \\ &= \sum_x P(X_0 = x)P(X_n = y \mid X_0 = x), \end{aligned}$$

we see that

$$(25) \qquad P(X_n = y) = \sum_x \pi_0(x)P^n(x, y).$$

This formula allows us to compute the distribution of X_n in terms of the initial distribution π_0 and the n-step transition function P^n.

For an alternative method of computing the distribution of X_n, observe that

$$\begin{aligned} P(X_{n+1} = y) &= \sum_x P(X_n = x, X_{n+1} = y) \\ &= \sum_x P(X_n = x)P(X_{n+1} = y \mid X_n = x), \end{aligned}$$

so that

$$P(X_{n+1} = y) = \sum_x P(X_n = x)P(x, y). \tag{26}$$

If we know the distribution of X_0, we can use (26) to find the distribution of X_1. Then, knowing the distribution of X_1, we can use (26) to find the distribution of X_2. Similarly, we can find the distribution of X_n by applying (26) n times.

We will use the notation $P_x(\ )$ to denote probabilities of various events defined in terms of a Markov chain starting at x. Thus

$$P_x(X_1 \neq a, X_2 \neq a, X_3 = a)$$

denotes the probability that a Markov chain starting at x is in a state a at time 3 but not at time 1 or at time 2. In terms of this notation, (19) can be rewritten as

$$\begin{aligned} P(X_{n+1} \in B_1, \ldots, X_{n+m} \in B_m \mid X_0 \in A_0, \ldots, X_{n-1} \in A_{n-1}, X_n = x) \\ = P_x(X_1 \in B_1, \ldots, X_m \in B_m). \end{aligned} \tag{27}$$

1.4.1. Hitting times. Let A be a subset of $\mathscr{S}$. The *hitting time* T_A of A is defined by

$$T_A = \min\,(n > 0: X_n \in A)$$

if $X_n \in A$ for some $n > 0$, and by $T_A = \infty$ if $X_n \notin A$ for all $n > 0$. In other words, T_A is the first positive time the Markov chain is in (*hits*) A. Hitting times play an important role in the theory of Markov chains. In this book we will be interested mainly in hitting times of sets consisting of a single point. We denote the hitting time of a point $a \in \mathscr{S}$ by T_a rather than by the more cumbersome notation $T_{\{a\}}$.

An important equation involving hitting times is given by

$$P^n(x, y) = \sum_{m=1}^{n} P_x(T_y = m)P^{n-m}(y, y), \qquad n \geq 1. \tag{28}$$

In order to verify (28) we note that the events $\{T_y = m, X_n = y\}$, $1 \leq m \leq n$, are disjoint and that

$$\{X_n = y\} = \bigcup_{m=1}^{n} \{T_y = m, X_n = y\}.$$

We have in effect decomposed the event $\{X_n = y\}$ according to the hitting time of y. We see from this decomposition that

$$\begin{aligned}
P^n(x, y) &= P_x(X_n = y) \\
&= \sum_{m=1}^{n} P_x(T_y = m, X_n = y) \\
&= \sum_{m=1}^{n} P_x(T_y = m)P(X_n = y \mid X_0 = x, T_y = m) \\
&= \sum_{m=1}^{n} P_x(T_y = m)P(X_n = y \mid X_0 = x, X_1 \neq y, \ldots, \\
&\qquad\qquad\qquad\qquad\qquad X_{m-1} \neq y, X_m = y) \\
&= \sum_{m=1}^{n} P_x(T_y = m)P^{n-m}(y, y),
\end{aligned}$$

and hence that (28) holds.

Example 8. Show that if a is an absorbing state, then $P^n(x, a) = P_x(T_a \leq n)$, $n \geq 1$.

If a is an absorbing state, then $P^{n-m}(a, a) = 1$ for $1 \leq m \leq n$, and hence (28) implies that

$$\begin{aligned}
P^n(x, a) &= \sum_{m=1}^{n} P_x(T_a = m)P^{n-m}(a, a) \\
&= \sum_{m=1}^{n} P_x(T_a = m) = P_x(T_a \leq n).
\end{aligned}$$

Observe that

$$P_x(T_y = 1) = P_x(X_1 = y) = P(x, y)$$

and that

$$P_x(T_y = 2) = \sum_{z \neq y} P_x(X_1 = z, X_2 = y) = \sum_{z \neq y} P(x, z)P(z, y).$$

For higher values of n the probabilities $P_x(T_y = n)$ can be found by using the formula

$$(29) \qquad P_x(T_y = n + 1) = \sum_{z \neq y} P(x, z)P_z(T_y = n), \qquad n \geq 1.$$

This formula is a consequence of (27), but it should also be directly obvious. For in order to go from x to y for the first time at time $n + 1$, it is necessary to go to some state $z \neq y$ at the first step and then go from z to y for the first time at the end of n additional steps.

1.4.2. Transition matrix. Suppose now that the state space $\mathscr{S}$ is finite, say $\mathscr{S} = \{0, 1, \ldots, d\}$. In this case we can think of P as the *transition matrix* having $d + 1$ rows and columns given by

$$\begin{array}{c c} & \begin{array}{ccc} 0 & \cdots & d \end{array} \\ \begin{array}{c} 0 \\ \vdots \\ d \end{array} & \begin{bmatrix} P(0, 0) & \cdots & P(0, d) \\ \vdots & & \vdots \\ P(d, 0) & \cdots & P(d, d) \end{bmatrix} \end{array}.$$

For example, the transition matrix of the gambler's ruin chain on $\{0, 1, 2, 3\}$ is

$$\begin{array}{c c} & \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 1 & 0 & 0 & 0 \\ q & 0 & p & 0 \\ 0 & q & 0 & p \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array}.$$

Similarly, we can regard P^n as an *n-step transition matrix.* Formula (24) with $m = n = 1$ becomes

$$P^2(x, y) = \sum_z P(x, z)P(z, y).$$

Recalling the definition of ordinary matrix multiplication, we observe that the two-step transition matrix P^2 is the product of the matrix P with itself. More generally, by setting $m = 1$ in (24) we see that

$$P^{n+1}(x, y) = \sum_z P^n(x, z)P(z, y). \tag{30}$$

It follows from (30) by induction that the n-step transition matrix P^n is the nth power of P.

An initial distribution π_0 can be thought of as a $(d + 1)$-dimensional row vector

$$\pi_0 = (\pi_0(0), \ldots, \pi_0(d)).$$

If we let π_n denote the $(d + 1)$-dimensional row vector

$$\pi_n = (P(X_n = 0), \ldots, P(X_n = d)),$$

then (25) and (26) can be written respectively as

$$\pi_n = \pi_0 P^n$$

and

$$\pi_{n+1} = \pi_n P.$$

The two-state Markov chain discussed in Section 1.1 is one of the few examples where P^n can be found very easily.

Example 9. Consider the two-state Markov chain having one-step transition matrix

$$P = \begin{bmatrix} 1 - p & p \\ q & 1 - q \end{bmatrix},$$

where $p + q > 0$. Find P^n.

In order to find $P^n(0, 0) = P_0(X_n = 0)$, we set $\pi_0(0) = 1$ in (3) and obtain

$$P^n(0, 0) = \frac{q}{p + q} + (1 - p - q)^n \frac{p}{p + q}.$$

In order to find $P^n(0, 1) = P_0(X_n = 1)$, we set $\pi_0(1) = 0$ in (4) and obtain

$$P^n(0, 1) = \frac{p}{p + q} - (1 - p - q)^n \frac{p}{p + q}.$$

Similarly, we conclude that

$$P(^n1, 0) = \frac{q}{p + q} - (1 - p - q)^n \frac{q}{p + q}$$

and

$$P^n(1, 1) = \frac{p}{p + q} + (1 - p - q)^n \frac{q}{p + q}.$$

It follows that

$$P^n = \frac{1}{p + q} \begin{bmatrix} q & p \\ q & p \end{bmatrix} + \frac{(1 - p - q)^n}{p + q} \begin{bmatrix} p & -p \\ -q & q \end{bmatrix}.$$

1.5. Transient and recurrent states

Let X_n, $n \geq 0$, be a Markov chain having state space $\mathscr{S}$ and transition function P. Set

$$\rho_{xy} = P_x(T_y < \infty).$$

Then ρ_{xy} denotes the probability that a Markov chain starting at x will be in state y at some positive time. In particular, ρ_{yy} denotes the probability that a Markov chain starting at y will ever return to y. A state y is called *recurrent* if $\rho_{yy} = 1$ and *transient* if $\rho_{yy} < 1$. If y is a recurrent state, a Markov chain starting at y returns to y with probability one. If y is a transient state, a Markov chain starting at y has positive probability $1 - \rho_{yy}$ of never returning to y. If y is an absorbing state, then $P_y(T_y = 1) =$

$P(y, y) = 1$ and hence $\rho_{yy} = 1$; thus an absorbing state is necessarily recurrent.

Let $1_y(z)$, $z \in \mathscr{S}$, denote the indicator function of the set $\{y\}$ defined by

$$1_y(z) = \begin{cases} 1, & z = y, \\ 0, & z \neq y. \end{cases}$$

Let $N(y)$ denote the number of times $n \geq 1$ that the chain is in state y. Since $1_y(X_n) = 1$ if the chain is in state y at time n and $1_y(X_n) = 0$ otherwise, we see that

$$N(y) = \sum_{n=1}^{\infty} 1_y(X_n). \tag{31}$$

The event $\{N(y) \geq 1\}$ is the same as the event $\{T_y < \infty\}$. Thus

$$P_x(N(y) \geq 1) = P_x(T_y < \infty) = \rho_{xy}.$$

Let m and n be positive integers. By (27), the probability that a Markov chain starting at x first visits y at time m and next visits y n units of time later is $P_x(T_y = m)P_y(T_y = n)$. Thus

$$\begin{aligned} P_x(N(y) \geq 2) &= \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} P_x(T_y = m)P_y(T_y = n) \\ &= \left(\sum_{m=1}^{\infty} P_x(T_y = m)\right)\left(\sum_{n=1}^{\infty} P_y(T_y = n)\right) \\ &= \rho_{xy}\rho_{yy}. \end{aligned}$$

Similarly we conclude that

$$P_x(N(y) \geq m) = \rho_{xy}\rho_{yy}^{m-1}, \qquad m \geq 1. \tag{32}$$

Since

$$P_x(N(y) = m) = P_x(N(y) \geq m) - P_x(N(y) \geq m + 1),$$

it follows from (32) that

$$P_x(N(y) = m) = \rho_{xy}\rho_{yy}^{m-1}(1 - \rho_{yy}), \qquad m \geq 1. \tag{33}$$

Also

$$P_x(N(y) = 0) = 1 - P_x(N(y) \geq 1),$$

so that

$$P_x(N(y) = 0) = 1 - \rho_{xy}. \tag{34}$$

These formulas are intuitively obvious. To see why (33) should be true, for example, observe that a chain starting at x visits state y exactly m times if and only if it visits y for a first time, returns to y $m - 1$ additional times, and then never again returns to y.

We use the notation $E_x(\ )$ to denote expectations of random variables defined in terms of a Markov chain starting at x. For example,

$$E_x(1_y(X_n)) = P_x(X_n = y) = P^n(x, y). \tag{35}$$

It follows from (31) and (35) that

$$\begin{aligned} E_x(N(y)) &= E_x\left(\sum_{n=1}^{\infty} 1_y(X_n)\right) \\ &= \sum_{n=1}^{\infty} E_x(1_y(X_n)) \\ &= \sum_{n=1}^{\infty} P^n(x, y). \end{aligned}$$

Set

$$G(x, y) = E_x(N(y)) = \sum_{n=1}^{\infty} P^n(x, y).$$

Then $G(x, y)$ denotes the expected number of visits to y for a Markov chain starting at x.

Theorem 1 (i) *Let y be a transient state. Then*

$$P_x(N(y) < \infty) = 1$$

and

$$G(x, y) = \frac{\rho_{xy}}{1 - \rho_{yy}}, \qquad x \in \mathscr{S}, \tag{36}$$

which is finite for all $x \in \mathscr{S}$.

(ii) *Let y be a recurrent state. Then $P_y(N(y) = \infty) = 1$ and $G(y, y) = \infty$. Also*

$$P_x(N(y) = \infty) = P_x(T_y < \infty) = \rho_{xy}, \qquad x \in \mathscr{S}. \tag{37}$$

If $\rho_{xy} = 0$, then $G(x, y) = 0$, while if $\rho_{xy} > 0$, then $G(x, y) = \infty$.

This theorem describes the fundamental difference between a transient state and a recurrent state. If y is a transient state, then no matter where the Markov chain starts, it makes only a finite number of visits to y and the expected number of visits to y is finite. Suppose instead that y is a recurrent state. Then if the Markov chain starts at y, it returns to y infinitely often. If the chain starts at some other state x, it may be impossible for it to ever hit y. If it is possible, however, and the chain does visit y at least once, then it does so infinitely often.

Proof. Let y be a transient state. Since $0 \leq \rho_{yy} < 1$, it follows from (32) that

$$P_x(N(y) = \infty) = \lim_{m\to\infty} P_x(N(y) \geq m) = \lim_{m\to\infty} \rho_{xy}\rho_{yy}^{m-1} = 0.$$

By (33)

$$\begin{aligned} G(x, y) &= E_x(N(y)) \\ &= \sum_{m=1}^{\infty} mP_x(N(y) = m) \\ &= \sum_{m=1}^{\infty} m\rho_{xy}\rho_{yy}^{m-1}(1 - \rho_{yy}). \end{aligned}$$

Substituting $t = \rho_{yy}$ in the power series

$$\sum_{m=1}^{\infty} mt^{m-1} = \frac{1}{(1-t)^2},$$

we conclude that

$$G(x, y) = \frac{\rho_{xy}}{1 - \rho_{yy}} < \infty.$$

This completes the proof of (i).

Now let y be recurrent. Then $\rho_{yy} = 1$ and it follows from (32) that

$$\begin{aligned} P_x(N(y) = \infty) &= \lim_{m\to\infty} P_x(N(y) \geq m) \\ &= \lim_{m\to\infty} \rho_{xy} = \rho_{xy}. \end{aligned}$$

In particular, $P_y(N(y) = \infty) = 1$. If a nonnegative random variable has positive probability of being infinite, its expectation is infinite. Thus

$$G(y, y) = E_y(N(y)) = \infty.$$

If $\rho_{xy} = 0$, then $P_x(T_y = m) = 0$ for all finite positive integers m, so (28) implies that $P^n(x, y) = 0$, $n \geq 1$; thus $G(x, y) = 0$ in this case. If $\rho_{xy} > 0$, then $P_x(N(y) = \infty) = \rho_{xy} > 0$ and hence

$$G(x, y) = E_x(N(y)) = \infty.$$

This completes the proof of Theorem 1. ∎

Let y be a transient state. Since

$$\sum_{n=1}^{\infty} P^n(x, y) = G(x, y) < \infty, \qquad x \in \mathscr{S},$$

we see that

$$\lim_{n\to\infty} P^n(x, y) = 0, \qquad x \in \mathscr{S}. \tag{38}$$

A Markov chain is called a *transient chain* if all of its states are transient and a *recurrent chain* if all of its states are recurrent. It is easy to see that a Markov chain having a finite state space must have at least one recurrent state and hence cannot possibly be a transient chain. For if $\mathscr{S}$ is finite and all states are transient, then by (38)

$$\begin{aligned} 0 &= \sum_{y \in \mathscr{S}} \lim_{n \to \infty} P^n(x, y) \\ &= \lim_{n \to \infty} \sum_{y \in \mathscr{S}} P^n(x, y) \\ &= \lim_{n \to \infty} P_x(X_n \in \mathscr{S}) \\ &= \lim_{n \to \infty} 1 = 1, \end{aligned}$$

which is a contradiction.

1.6. Decomposition of the state space

Let x and y be two not necessarily distinct states. We say that x *leads to* y if $\rho_{xy} > 0$. It is left as an exercise for the reader to show that x leads to y if and only if $P^n(x, y) > 0$ for some positive integer n. It is also left to the reader to show that if x leads to y and y leads to z, then x leads to z.

Theorem 2 *Let x be a recurrent state and suppose that x leads to y. Then y is recurrent and $\rho_{xy} = \rho_{yx} = 1$.*

Proof. We assume that $y \neq x$, for otherwise there is nothing to prove. Since

$$P_x(T_y < \infty) = \rho_{xy} > 0,$$

we see that $P_x(T_y = n) > 0$ for some positive integer n. Let n_0 be the least such positive integer, i.e., set

$$n_0 = \min(n \geq 1 : P_x(T_y = n) > 0). \tag{39}$$

It follows easily from (39) and (28) that $P^{n_0}(x, y) > 0$ and

$$P^m(x, y) = 0, \qquad 1 \leq m < n_0. \tag{40}$$

Since $P^{n_0}(x, y) > 0$, we can find states $y_1, \ldots, y_{n_0 - 1}$ such that

$$P_x(X_1 = y_1, \ldots, X_{n_0 - 1} = y_{n_0 - 1}, X_{n_0} = y) = P(x, y_1) \cdots P(y_{n_0 - 1}, y) > 0.$$

None of the states $y_1, \ldots, y_{n_0 - 1}$ equals x or y; for if one of them did equal x or y, it would be possible to go from x to y with positive probability in fewer than n_0 steps, in contradiction to (40).

We will now show that $\rho_{yx} = 1$. Suppose on the contrary that $\rho_{yx} < 1$. Then a Markov chain starting at y has positive probability $1 - \rho_{yx}$ of never hitting x. More to the point, a Markov chain starting at x has the positive probability

$$P(x, y_1) \cdots P(y_{n_0-1}, y)(1 - \rho_{yx})$$

of visiting the states $y_1, \ldots, y_{n_0-1}, y$ successively in the first n_0 times and never returning to x after time n_0. But if this happens, the Markov chain never returns to x at any time $n \geq 1$, so we have contradicted the assumption that x is a recurrent state.

Since $\rho_{yx} = 1$, there is a positive integer n_1 such that $P^{n_1}(y, x) > 0$. Now

$$\begin{aligned} P^{n_1+n+n_0}(y, y) &= P_y(X_{n_1+n+n_0} = y) \\ &\geq P_y(X_{n_1} = x, X_{n_1+n} = x, X_{n_1+n+n_0} = y) \\ &= P^{n_1}(y, x)P^n(x, x)P^{n_0}(x, y). \end{aligned}$$

Hence

$$\begin{aligned} G(y, y) &\geq \sum_{n=n_1+1+n_0}^{\infty} P^n(y, y) \\ &= \sum_{n=1}^{\infty} P^{n_1+n+n_0}(y, y) \\ &\geq P^{n_1}(y, x)P^{n_0}(x, y) \sum_{n=1}^{\infty} P^n(x, x) \\ &= P^{n_1}(y, x)P^{n_0}(x, y)G(x, x) = +\infty, \end{aligned}$$

from which it follows that y is also a recurrent state.

Since y is recurrent and y leads to x, we see from the part of the theorem that has already been verified that $\rho_{xy} = 1$. This completes the proof. ∎

A nonempty set C of states is said to be *closed* if no state inside of C leads to any state outside of C, i.e., if

(41) $$\rho_{xy} = 0, \qquad x \in C \text{ and } y \notin C.$$

Equivalently (see Exercise 16), C is closed if and only if

(42) $$P^n(x, y) = 0, \qquad x \in C,\ y \notin C, \text{ and } n \geq 1.$$

Actually, even from the weaker condition

(43) $$P(x, y) = 0, \qquad x \in C \text{ and } y \notin C,$$

we can prove that C is closed. For if (43) holds, then for $x \in C$ and $y \notin C$

$$\begin{aligned} P^2(x, y) &= \sum_{z \in \mathscr{S}} P(x, z)P(z, y) \\ &= \sum_{z \in C} P(x, z)P(z, y) = 0, \end{aligned}$$

and (42) follows by induction. If C is closed, then a Markov chain starting in C will, with probability one, stay in C for all time. If a is an absorbing state, then $\{a\}$ is closed.

A closed set C is called *irreducible* if x leads to y for all choices of x and y in C. It follows from Theorem 2 that if C is an irreducible closed set, then either every state in C is recurrent or every state in C is transient. The next result is an immediate consequence of Theorems 1 and 2.

Corollary 1 *Let C be an irreducible closed set of recurrent states. Then $\rho_{xy} = 1$, $P_x(N(y) = \infty) = 1$, and $G(x, y) = \infty$ for all choices of x and y in C.*

An *irreducible Markov chain* is a chain whose state space is irreducible, that is, a chain in which every state leads back to itself and also to every other state. Such a Markov chain is necessarily either a transient chain or a recurrent chain. Corollary 1 implies, in particular, that an irreducible recurrent Markov chain visits every state infinitely often with probability one.

We saw in Section 1.5 that if $\mathscr{S}$ is finite, it contains at least one recurrent state. The same argument shows that any finite closed set of states contains at least one recurrent state. Now let C be a finite irreducible closed set. We have seen that either every state in C is transient or every state in C is recurrent, and that C has at least one recurrent state. It follows that every state in C is recurrent. We summarize this result:

Theorem 3 *Let C be a finite irreducible closed set of states. Then every state in C is recurrent.*

Consider a Markov chain having a finite number of states. Theorem 3 implies that if the chain is irreducible it must be recurrent. If the chain is not irreducible, we can use Theorems 2 and 3 to determine which states are recurrent and which are transient.

Example 10. Consider a Markov chain having the transition matrix

$$\begin{array}{c c}
 & \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & 5 \end{array} \\
\begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} &
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
\frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 & 0 \\
0 & \frac{1}{5} & \frac{2}{5} & \frac{1}{5} & 0 & \frac{1}{5} \\
0 & 0 & 0 & \frac{1}{6} & \frac{1}{3} & \frac{1}{2} \\
0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\
0 & 0 & 0 & \frac{1}{4} & 0 & \frac{3}{4}
\end{bmatrix}
\end{array}.$$

Determine which states are recurrent and which states are transient.

As a first step in studying this Markov chain, we determine by inspection which states lead to which other states. This can be indicated in matrix form as

$$\begin{array}{c@{\;}c} & \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} & \begin{bmatrix} + & 0 & 0 & 0 & 0 & 0 \\ + & + & + & + & + & + \\ + & + & + & + & + & + \\ 0 & 0 & 0 & + & + & + \\ 0 & 0 & 0 & + & + & + \\ 0 & 0 & 0 & + & + & + \end{bmatrix} \end{array}.$$

The x, y element of this matrix is $+$ or 0 according as ρ_{xy} is positive or zero, i.e., according as x does or does not lead to y. Of course, if $P(x, y) > 0$, then $\rho_{xy} > 0$. The converse is certainly not true in general. For example, $P(2, 0) = 0$; but

$$P^2(2, 0) = P(2, 1)P(1, 0) = \tfrac{1}{5} \cdot \tfrac{1}{4} = \tfrac{1}{20} > 0,$$

so that $\rho_{20} > 0$.

State 0 is an absorbing state, and hence also a recurrent state. We see clearly from the matrix of $+$'s and 0's that $\{3, 4, 5\}$ is an irreducible closed set. Theorem 3 now implies that 3, 4, and 5 are recurrent states. States 1 and 2 both lead to 0, but neither can be reached from 0. We see from Theorem 2 that 1 and 2 must both be transient states. In summary, states 1 and 2 are transient, and states 0, 3, 4, and 5 are recurrent.

Let $\mathscr{S}_T$ denote the collection of transient states in $\mathscr{S}$, and let $\mathscr{S}_R$ denote the collection of recurrent states in $\mathscr{S}$. In Example 10, $\mathscr{S}_T = \{1, 2\}$ and $\mathscr{S}_R = \{0, 3, 4, 5\}$. The set $\mathscr{S}_R$ can be decomposed into the disjoint irreducible closed sets $C_1 = \{0\}$ and $C_2 = \{3, 4, 5\}$. The next theorem shows that such a decomposition is always possible whenever $\mathscr{S}_R$ is nonempty.

Theorem 4 *Suppose that the set $\mathscr{S}_R$ of recurrent states is nonempty. Then $\mathscr{S}_R$ is the union of a finite or countably infinite number of disjoint irreducible closed sets $C_1, C_2, \ldots$.*

Proof. Choose $x \in \mathscr{S}_R$ and let C be the set of all states y in $\mathscr{S}_R$ such that x leads to y. Since x is recurrent, $\rho_{xx} = 1$ and hence $x \in C$. We will now verify that C is an irreducible closed set. Suppose that y is in C and y leads to z. Since y is recurrent, it follows from Theorem 2 that z is recurrent. Since x leads to y and y leads to z, we conclude that x leads to z. Thus z is in C. This shows that C is closed. Suppose that y and z are both in C. Since x is recurrent and x leads to y, it follows from

Theorem 2 that y leads to x. Since y leads to x and x leads to z, we conclude that y leads to z. This shows that C is irreducible.

To complete the proof of the theorem, we need only show that if C and D are two irreducible closed subsets of $\mathscr{S}_R$, they are either disjoint or identical. Suppose they are not disjoint and let x be in both C and D. Choose y in C. Now x leads to y, since x is in C and C is irreducible. Since D is closed, x is in D, and x leads to y, we conclude that y is in D. Thus every state in C is also in D. Similarly every state in D is also in C, so that C and D are identical. ∎

We can use our decomposition of the state space of a Markov chain to understand the behavior of such a system. If the Markov chain starts out in one of the irreducible closed sets C_i of recurrent states, it stays in C_i forever and, with probability one, visits every state in C_i infinitely often. If the Markov chain starts out in the set of transient states $\mathscr{S}_T$, it either stays in $\mathscr{S}_T$ forever or, at some time, enters one of the sets C_i and stays there from that time on, again visiting every state in that C_i infinitely often.

1.6.1 Absorption probabilities. Let C be one of the irreducible closed sets of recurrent states, and let $\rho_C(x) = P_x(T_C < \infty)$ be the probability that a Markov chain starting at x eventually hits C. Since the chain remains permanently in C once it hits that set, we call $\rho_C(x)$ the probability that a chain starting at x is *absorbed* by the set C. Clearly $\rho_C(x) = 1$, $x \in C$, and $\rho_C(x) = 0$ if x is a recurrent state not in C. It is not so clear how to compute $\rho_C(x)$ for $x \in \mathscr{S}_T$, the set of transient states.

If there are only a finite number of transient states, and in particular if $\mathscr{S}$ itself is finite, it is always possible to compute $\rho_C(x)$, $x \in \mathscr{S}_T$, by solving a system of linear equations in which there are as many equations as unknowns, i.e., members of $\mathscr{S}_T$. To understand why this is the case, observe that if $x \in \mathscr{S}_T$, a chain starting at x can enter C only by entering C at time 1 or by being in $\mathscr{S}_T$ at time 1 and entering C at some future time. The former event has probability $\sum_{y \in C} P(x, y)$ and the latter event has probability $\sum_{y \in \mathscr{S}_T} P(x, y)\rho_C(y)$. Thus

$$(44) \qquad \rho_C(x) = \sum_{y \in C} P(x, y) + \sum_{y \in \mathscr{S}_T} P(x, y)\rho_C(y), \qquad x \in \mathscr{S}_T.$$

Equation (44) holds whether $\mathscr{S}_T$ is finite or infinite, but it is far from clear how to solve (44) for the unknowns $\rho_C(x)$, $x \in \mathscr{S}_T$, when $\mathscr{S}_T$ is infinite. An additional difficulty is that if $\mathscr{S}_T$ is infinite, then (44) need not have a unique solution. Fortunately this difficulty does not arise if $\mathscr{S}_T$ is finite.

Theorem 5 *Suppose the set $\mathscr{S}_T$ of transient states is finite and let C be an irreducible closed set of recurrent states. Then the system of equations*

$$f(x) = \sum_{y \in C} P(x, y) + \sum_{y \in \mathscr{S}_T} P(x, y)f(y), \qquad x \in \mathscr{S}_T, \tag{45}$$

has the unique solution

$$f(x) = \rho_C(x), \qquad x \in \mathscr{S}_T. \tag{46}$$

Proof. If (45) holds, then

$$f(y) = \sum_{z \in C} P(y, z) + \sum_{z \in \mathscr{S}_T} P(y, z)f(z), \qquad y \in \mathscr{S}_T.$$

Substituting this into (45) we find that

$$f(x) = \sum_{y \in C} P(x, y) + \sum_{y \in \mathscr{S}_T} \sum_{z \in C} P(x, y)P(y, z)$$
$$+ \sum_{y \in \mathscr{S}_T} \sum_{z \in \mathscr{S}_T} P(x, y)P(y, z)f(z).$$

The sum of the first two terms is just $P_x(T_C \leq 2)$, and the third term reduces to $\sum_{z \in \mathscr{S}_T} P^2(x, z)f(z)$, which is the same as $\sum_{y \in \mathscr{S}_T} P^2(x, y)f(y)$. Thus

$$f(x) = P_x(T_C \leq 2) + \sum_{y \in \mathscr{S}_T} P^2(x, y)f(y).$$

By repeating this argument indefinitely or by using induction, we conclude that for all positive integers n

$$f(x) = P_x(T_C \leq n) + \sum_{y \in \mathscr{S}_T} P^n(x, y)f(y), \qquad x \in \mathscr{S}_T. \tag{47}$$

Since each $y \in \mathscr{S}_T$ is transient, it follows from (38) that

$$\lim_{n \to \infty} P^n(x, y) = 0, \qquad x \in \mathscr{S} \quad \text{and} \quad y \in \mathscr{S}_T. \tag{48}$$

According to the assumptions of the theorem, $\mathscr{S}_T$ is a finite set. It therefore follows from (48) that the sum in (47) approaches zero as $n \to \infty$. Consequently for $x \in \mathscr{S}_T$

$$f(x) = \lim_{n \to \infty} P_x(T_C \leq n) = P_x(T_C < \infty) = \rho_C(x),$$

as desired. ∎

Example 11. Consider the Markov chain discussed in Example 10. Find

$$\rho_{10} = \rho_{\{0\}}(1) \qquad \text{and} \qquad \rho_{20} = \rho_{\{0\}}(2).$$

From (44) and the transition matrix in Example 10, we see that ρ_{10} and ρ_{20} are determined by the equations

$$\rho_{10} = \tfrac{1}{4} + \tfrac{1}{2}\rho_{10} + \tfrac{1}{4}\rho_{20}$$

and

$$\rho_{20} = \tfrac{1}{5}\rho_{10} + \tfrac{2}{5}\rho_{20}.$$

Solving these equations we find that $\rho_{10} = \frac{3}{5}$ and $\rho_{20} = \frac{1}{5}$.

By similar methods we conclude that $\rho_{\{3,4,5\}}(1) = \frac{2}{5}$ and $\rho_{\{3,4,5\}}(2) = \frac{4}{5}$. Alternatively, we can obtain these probabilities by subtracting $\rho_{\{0\}}(1)$ and $\rho_{\{0\}}(2)$ from 1, since if there are only a finite number of transient states,

$$(49) \qquad \sum_i \rho_{C_i}(x) = 1, \qquad x \in \mathscr{S}_T.$$

To verify (49) we note that for $x \in \mathscr{S}_T$

$$\sum_i \rho_{C_i}(x) = \sum_i P_x(T_{C_i} < \infty) = P_x(T_{\mathscr{S}_R} < \infty).$$

Since there are only a finite number of transient states and each transient state is visited only finitely many times, the probability $P_x(T_{\mathscr{S}_R} < \infty)$ that a recurrent state will eventually be hit is 1, so (49) holds.

Once a Markov chain starting at a transient state x enters an irreducible closed set C of recurrent states, it visits every state in C. Thus

$$(50) \qquad \rho_{xy} = \rho_C(x), \qquad x \in \mathscr{S}_T \text{ and } y \in C.$$

It follows from (50) that in our previous example

$$\rho_{13} = \rho_{14} = \rho_{15} = \rho_{\{3,4,5\}}(1) = \tfrac{2}{5}$$

and

$$\rho_{23} = \rho_{24} = \rho_{25} = \rho_{\{3,4,5\}}(2) = \tfrac{4}{5}.$$

1.6.2. Martingales. Consider a Markov chain having state space $\{0, \ldots, d\}$ and transition function P such that

$$(51) \qquad \sum_{y=0}^{d} yP(x, y) = x, \qquad x = 0, \ldots, d.$$

Now

$$\begin{aligned} E[X_{n+1} \mid X_0 &= x_0, \ldots, X_{n-1} = x_{n-1}, X_n = x] \\ &= \sum_{y=0}^{d} yP[X_{n+1} = y \mid X_0 = x_0, \ldots, X_{n-1} = x_{n-1}, X_n = x] \\ &= \sum_{y=0}^{d} yP(x, y) \end{aligned}$$

by the Markov property. We conclude from (51) that

$$(52) \qquad E[X_{n+1} \mid X_0 = x_0, \ldots, X_{n-1} = x_{n-1}, X_n = x] = x,$$

i.e., that the expected value of X_{n+1} given the past and present values of $X_0, \ldots, X_n$ equals the present value of X_n. A sequence of random variables having this property is called a *martingale*. Martingales, which need not be Markov chains, play a very important role in modern probability theory. They arose first in connection with gambling. If X_n denotes the capital of a gambler after time n and if all bets are "fair," that is, if they result in zero expected gain to the gambler, then X_n, $n \geq 0$, forms a martingale. Gamblers were naturally interested in finding some betting strategy, such as increasing their bets until they win, that would give them a net expected gain after making a series of fair bets. That this has been shown to be mathematically impossible does not seem to have deterred them from their quest.

It follows from (51) that

$$\sum_{y=0}^{d} yP(0, y) = 0,$$

and hence that $P(0, 1) = \cdots = P(0, d) = 0$. Thus 0 is necessarily an absorbing state. It follows similarly that d is an absorbing state. Consider now a Markov chain satisfying (51) and having no absorbing states other than 0 and d. It is left as an exercise for the reader to show that under these conditions the states $1, \ldots, d - 1$ each lead to state 0, and hence each is a transient state. If the Markov chain starts at x, it will eventually enter one of the two absorbing states 0 and d and remain there permanently.

It follows from Example 8 that

$$\begin{aligned} E_x(X_n) &= \sum_{y=0}^{d} yP_x(X_n = y) \\ &= \sum_{y=0}^{d} yP^n(x, y) \\ &= \sum_{y=1}^{d-1} yP^n(x, y) + dP^n(x, d) \\ &= \sum_{y=1}^{d-1} yP^n(x, y) + dP_x(T_d \leq n). \end{aligned}$$

Since states $1, 2, \ldots, d - 1$ are transient, we see that $P^n(x, y) \to 0$ as $n \to \infty$ for $y = 1, 2, \ldots, d - 1$. Consequently,

$$\lim_{n \to \infty} E_x(X_n) = dP_x(T_d < \infty) = d\rho_{xd}.$$

On the other hand, it follows from (51) (see Exercise 13(a)) that $EX_n = EX_{n-1} = \cdots = EX_0$ and hence that $E_x(X_n) = x$. Thus

$$\lim_{n \to \infty} E_x(X_n) = x.$$

By equating the two values of this limit, we conclude that

$$\rho_{xd} = \frac{x}{d}, \qquad x = 0, \ldots, d. \tag{53}$$

Since $\rho_{x0} + \rho_{xd} = 1$, it follows from (53) that

$$\rho_{x0} = 1 - \frac{x}{d}, \qquad x = 0, \ldots, d.$$

Of course, once (53) is conjectured, it is easily proved directly from Theorem 5. We need only verify that for $x = 1, \ldots, d - 1$,

$$\frac{x}{d} = P(x, d) + \sum_{y=1}^{d-1} \frac{y}{d} P(x, y). \tag{54}$$

Clearly (54) follows from (51).

The genetics chain introduced in Example 7 satisfies (51) as does a gambler's ruin chain on $\{0, 1, \ldots, d\}$ having transition matrix of the form

$$\begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & & & & \cdot \\ \cdot & \frac{1}{2} & 0 & \frac{1}{2} & & & \cdot \\ \cdot & & & & & & \cdot \\ \cdot & & & & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & 1 \end{bmatrix}.$$

Suppose two gamblers make a series of one dollar bets until one of them goes broke, and suppose that each gambler has probability $\frac{1}{2}$ of winning any given bet. If the first gambler has an initial capital of x dollars and the second gambler has an initial capital of $d - x$ dollars, then the second gambler has probability $\rho_{xd} = x/d$ of going broke and the first gambler has probability $1 - (x/d)$ of going broke.

1.7. Birth and death chains

For an irreducible Markov chain either every state is recurrent or every state is transient, so that an irreducible Markov chain is either a recurrent chain or a transient chain. An irreducible Markov chain having only finitely many states is necessarily recurrent. It is generally difficult to decide whether an irreducible chain having infinitely many states is recurrent or transient. We are able to do so, however, for the birth and death chain.

Consider a birth and death chain on the nonnegative integers or on the finite set $\{0, \ldots, d\}$. In the former case we set $d = \infty$. The transition function is of the form

$$P(x, y) = \begin{cases} q_x, & y = x - 1, \\ r_x, & y = x, \\ p_x, & y = x + 1, \end{cases}$$

where $p_x + q_x + r_x = 1$ for $x \in \mathscr{S}$, $q_0 = 0$, and $p_d = 0$ if $d < \infty$. We assume additionally that p_x and q_x are positive for $0 < x < d$.

For a and b in $\mathscr{S}$ such that $a < b$, set

$$u(x) = P_x(T_a < T_b), \qquad a < x < b,$$

and set $u(a) = 1$ and $u(b) = 0$. If the birth and death chain starts at y, then in one step it goes to $y - 1$, y, or $y + 1$ with respective probabilities q_y, r_y, or p_y. It follows that

$$(55) \quad u(y) = q_y u(y - 1) + r_y u(y) + p_y u(y + 1), \qquad a < y < b.$$

Since $r_y = 1 - p_y - q_y$, we can rewrite (55) as

$$(56) \qquad u(y + 1) - u(y) = \frac{q_y}{p_y}(u(y) - u(y - 1)), \qquad a < y < b.$$

Set $\gamma_0 = 1$ and

$$(57) \qquad \gamma_y = \frac{q_1 \cdots q_y}{p_1 \cdots p_y}, \qquad 0 < y < d.$$

From (56) we see that

$$u(y + 1) - u(y) = \frac{\gamma_y}{\gamma_{y-1}}(u(y) - u(y - 1)), \qquad a < y < b,$$

from which it follows that

$$\begin{aligned} u(y + 1) - u(y) &= \frac{\gamma_{a+1}}{\gamma_a} \cdots \frac{\gamma_y}{\gamma_{y-1}}(u(a + 1) - u(a)) \\ &= \frac{\gamma_y}{\gamma_a}(u(a + 1) - u(a)). \end{aligned}$$

Consequently,

$$(58) \quad u(y) - u(y + 1) = \frac{\gamma_y}{\gamma_a}(u(a) - u(a + 1)), \qquad a \le y < b.$$

Summing (58) on $y = a, \ldots, b - 1$ and recalling that $u(a) = 1$ and $u(b) = 0$, we conclude that

$$\frac{u(a) - u(a + 1)}{\gamma_a} = \frac{1}{\sum_{y=a}^{b-1} \gamma_y}.$$

Thus (58) becomes

$$u(y) - u(y+1) = \frac{\gamma_y}{\sum_{y=a}^{b-1} \gamma_y}, \qquad a \le y < b.$$

Summing this equation on $y = x, \ldots, b - 1$ and again using the formula $u(b) = 0$, we obtain

$$u(x) = \frac{\sum_{y=x}^{b-1} \gamma_y}{\sum_{y=a}^{b-1} \gamma_y}, \qquad a < x < b.$$

It now follows from the definition of $u(x)$ that

$$P_x(T_a < T_b) = \frac{\sum_{y=x}^{b-1} \gamma_y}{\sum_{y=a}^{b-1} \gamma_y}, \qquad a < x < b. \tag{59}$$

By subtracting both sides of (59) from 1, we see that

$$P_x(T_b < T_a) = \frac{\sum_{y=a}^{x-1} \gamma_y}{\sum_{y=a}^{b-1} \gamma_y}, \qquad a < x < b. \tag{60}$$

Example 12. A gambler playing roulette makes a series of one dollar bets. He has respective probabilities 9/19 and 10/19 of winning and losing each bet. The gambler decides to quit playing as soon as his net winnings reach 25 dollars or his net losses reach 10 dollars.

(a) Find the probability that when he quits playing he will have won 25 dollars.

(b) Find his expected loss.

The problem fits into our scheme if we let X_n denote the capital of the gambler at time n with $X_0 = 10$. Then X_n, $n \ge 0$, forms a birth and death chain on $\{0, 1, \ldots, 35\}$ with birth and death rates

$$p_x = 9/19, \qquad 0 < x < 35,$$

and

$$q_x = 10/19, \qquad 0 < x < 35.$$

States 0 and 35 are absorbing states. Formula (60) is applicable with $a = 0$, $x = 10$, and $b = 35$. We conclude that

$$\gamma_y = (10/9)^y, \qquad 0 \le y \le 34,$$

and hence that

$$P_{10}(T_{35} < T_0) = \frac{\sum_{y=0}^{9} (10/9)^y}{\sum_{y=0}^{34} (10/9)^y} = \frac{(10/9)^{10} - 1}{(10/9)^{35} - 1} = .047.$$

Thus the gambler has probability .047 of winning 25 dollars. His expected loss in dollars is $10 - 35(.047)$, which equals \$8.36.

In the remainder of this section we consider a birth and death chain on the nonnegative integers which is irreducible, i.e., such that $p_x > 0$ for $x \geq 0$ and $q_x > 0$ for $x \geq 1$. We will determine when such a chain is recurrent and when it is transient.

As a special case of (59),

$$P_1(T_0 < T_n) = 1 - \frac{1}{\sum_{y=0}^{n-1} \gamma_y}, \qquad n > 1. \tag{61}$$

Consider now a birth and death chain starting in state 1. Since the birth and death chain can move at most one step to the right at a time (considering the transition from state to state as movement along the real number line),

$$1 \leq T_2 < T_3 < \cdots. \tag{62}$$

It follows from (62) that $\{T_0 < T_n\}$, $n > 1$, forms a nondecreasing sequence of events. We conclude from Theorem 1 of Chapter 1 of Volume I[1] that

$$\lim_{n \to \infty} P_1(T_0 < T_n) = P_1(T_0 < T_n \text{ for some } n > 1). \tag{63}$$

Equation (62) implies that $T_n \geq n$ and thus $T_n \to \infty$ as $n \to \infty$; hence the event $\{T_0 < T_n \text{ for some } n > 1\}$ occurs if and only if the event $\{T_0 < \infty\}$ occurs. We can therefore rewrite (63) as

$$\lim_{n \to \infty} P_1(T_0 < T_n) = P_1(T_0 < \infty). \tag{64}$$

It follows from (61) and (64) that

$$P_1(T_0 < \infty) = 1 - \frac{1}{\sum_{y=0}^{\infty} \gamma_y}. \tag{65}$$

We are now in position to show that the birth and death chain is recurrent if and only if

$$\sum_{y=0}^{\infty} \gamma_y = \infty. \tag{66}$$

If the birth and death chain is recurrent, then $P_1(T_0 < \infty) = 1$ and (66) follows from (65). To obtain the converse, we observe that $P(0, y) = 0$ for $y \geq 2$, and hence

$$P_0(T_0 < \infty) = P(0, 0) + P(0, 1)P_1(T_0 < \infty). \tag{67}$$

---

[1] Paul G. Hoel, Sidney C. Port, and Charles J. Stone, ***Introduction to Probability Theory*** (Boston: Houghton Mifflin Co., 1971), p. 13.

Suppose (66) holds. Then by (65)

$$P_1(T_0 < \infty) = 1.$$

From this and (67) we conclude that

$$P_0(T_0 < \infty) = P(0, 0) + P(0, 1) = 1.$$

Thus 0 is a recurrent state, and since the chain is assumed to be irreducible, it must be a recurrent chain.

In summary, we have shown that an irreducible birth and death chain on $\{0, 1, 2, \ldots\}$ is recurrent if and only if

$$\sum_{x=1}^{\infty} \frac{q_1 \cdots q_x}{p_1 \cdots p_x} = \infty. \tag{68}$$

Example 13. Consider the birth and death chain on $\{0, 1, 2, \ldots\}$ defined by

$$p_x = \frac{x + 2}{2(x + 1)} \quad \text{and} \quad q_x = \frac{x}{2(x + 1)}, \quad x \geq 0.$$

Determine whether this chain is recurrent or transient.

Since

$$\frac{q_x}{p_x} = \frac{x}{x + 2},$$

it follows that

$$\begin{aligned}
\gamma_x &= \frac{q_1 \cdots q_x}{p_1 \cdots p_x} = \frac{1 \cdot 2 \cdots x}{3 \cdot 4 \cdots (x + 2)} \\
&= \frac{2}{(x + 1)(x + 2)} = 2\left(\frac{1}{x + 1} - \frac{1}{x + 2}\right).
\end{aligned}$$

Thus

$$\begin{aligned}
\sum_{x=1}^{\infty} \gamma_x &= 2 \sum_{x=1}^{\infty} \left(\frac{1}{x + 1} - \frac{1}{x + 2}\right) \\
&= 2(\tfrac{1}{2} - \tfrac{1}{3} + \tfrac{1}{3} - \tfrac{1}{4} + \tfrac{1}{4} - \tfrac{1}{5} + \cdots) \\
&= 2 \cdot \tfrac{1}{2} = 1.
\end{aligned}$$

We conclude that the chain is transient.

1.8. Branching and queuing chains

In this section we will describe which branching chains are certain of extinction and which are not. We will also describe which queuing chains are transient and which are recurrent. The proofs of these results are somewhat complicated and will be given in the appendix to this chapter. These proofs can be skipped with no loss of continuity. It is interesting to note that the proofs of the results for the branching chain and the queuing chain are very similar, whereas the results themselves appear quite dissimilar.

1.8.1. Branching chain. Consider the branching chain introduced in Example 6. The extinction probability ρ of the chain is the probability that the descendants of a given particle eventually become extinct. Clearly

$$\rho = \rho_{10} = P_1(T_0 < \infty).$$

Suppose there are x particles present initially. Since the numbers of offspring of these particles in the various generations are chosen independently of each other, the probability ρ_{x0} that the descendants of each of the x particles eventually become extinct is just the xth power of the probability that the descendants of any one particle eventually become extinct. In other words,

$$\rho_{x0} = \rho^x, \qquad x = 1, 2, \ldots. \tag{69}$$

Recall from Example 6 that a particle gives rise to ξ particles in the next generation, where ξ is a random variable having density f. If $f(1) = 1$, the branching chain is degenerate in that every state is an absorbing state. Thus we suppose that $f(1) < 1$. Then state 0 is an absorbing state. It is left as an exercise for the reader to show that every state other than 0 is transient. From this it follows that, with probability one, the branching chain is either absorbed at 0 or approaches $+\infty$. We conclude from (69) that

$$P_x(\lim_{n\to\infty} X_n = \infty) = 1 - \rho^x, \qquad x = 1, 2, \ldots.$$

Clearly it is worthwhile to determine ρ or at least to determine when $\rho = 1$ and when $\rho < 1$. This can be done using arguments based upon the formula

$$\Phi(\rho) = \rho, \tag{70}$$

where Φ is the probability generating function of f, defined by

$$\Phi(t) = f(0) + \sum_{y=1}^{\infty} f(y)t^y, \qquad 0 \le t \le 1.$$

To verify (70) we observe that (see Exercise 9(b))

$$\begin{aligned}\rho = \rho_{10} &= P(1, 0) + \sum_{y=1}^{\infty} P(1, y)\rho_{y0} \\ &= P(1, 0) + \sum_{y=1}^{\infty} P(1, y)\rho^y \\ &= f(0) + \sum_{y=1}^{\infty} f(y)\rho^y \\ &= \Phi(\rho).\end{aligned}$$

Let μ denote the expected number of offspring of any given particle. Suppose $\mu \leq 1$. Then the equation $\Phi(t) = t$ has no roots in $[0, 1)$ (under our assumption that $f(1) < 1$), and hence $\rho = 1$. Thus *ultimate extinction is certain if* $\mu \leq 1$ *and* $f(1) < 1$.

Suppose instead that $\mu > 1$. Then the equation $\Phi(t) = t$ has a unique root ρ_0 in $[0, 1)$, and hence ρ equals either ρ_0 or 1. Actually ρ always equals ρ_0. Consequently, *if* $\mu > 1$ *the probability of ultimate extinction is less than one.*

The proofs of these results will be given in the appendix. The results themselves are intuitively very reasonable. If $\mu < 1$, then on the average each particle gives rise to fewer than one new particle, so we would expect the population to die out eventually. If $\mu > 1$, then on the average each particle gives rise to more than one new particle. In this case we would expect that the population has positive probability of growing rapidly, indeed geometrically fast, as time goes on. The case $\mu = 1$ is borderline; but since $\rho = 1$ when $\mu < 1$, it is plausible by "continuity" that $\rho = 1$ also when $\mu = 1$.

Example 14. Suppose that every man in a certain society has exactly three children, which independently have probability one-half of being a boy and one-half of being a girl. Suppose also that the number of males in the nth generation forms a branching chain. Find the probability that the male line of a given man eventually becomes extinct.

The density f of the number of male children of a given man is the binomial density with parameters $n = 3$ and $p = \frac{1}{2}$. Thus $f(0) = \frac{1}{8}$, $f(1) = \frac{3}{8}$, $f(2) = \frac{3}{8}$, $f(3) = \frac{1}{8}$, and $f(x) = 0$ for $x \geq 4$. The mean number of male offspring is $\mu = \frac{3}{2}$. Since $\mu > 1$, the extinction probability ρ is the root of the equation

$$\frac{1}{8} + \frac{3}{8}t + \frac{3}{8}t^2 + \frac{1}{8}t^3 = t$$

lying in $[0, 1)$. We can rewrite this equation as

$$t^3 + 3t^2 - 5t + 1 = 0,$$

or equivalently as

$$(t - 1)(t^2 + 4t - 1) = 0.$$

This equation has three roots, namely, 1, $-\sqrt{5} - 2$, and $\sqrt{5} - 2$. Consequently, $\rho = \sqrt{5} - 2$.

1.8.2. Queuing chain.

Consider the queuing chain introduced in Example 5. Let $\xi_1, \xi_2, \ldots$ and μ be as in that example. In this section we will indicate when the queuing chain is recurrent and when it is transient.

Let μ denote the expected number of customers arriving in unit time. Suppose first that $\mu > 1$. Since at most one person is served at a time and on the average more than one new customer enters the queue at a time, it would appear that as time goes on more and more people will be waiting for service and that the queue length will approach infinity. This is indeed the case, so that *if $\mu > 1$ the queuing chain is transient.*

In discussing the case $\mu \leq 1$, we will assume that the chain is irreducible (see Exercises 37 and 38 for necessary and sufficient conditions for irreducibility and for results when the queuing chain is not irreducible). Suppose first that $\mu < 1$. Then on the average fewer than one new customer will enter the queue in unit time. Since one customer is served whenever the queue is nonempty, we would expect that, regardless of the initial length of the queue, it will become empty at some future time. This is indeed the case and, in particular, 0 is a recurrent state. The case $\mu = 1$ is borderline, but again it turns out that 0 is a recurrent state. Thus *if $\mu \leq 1$ and the queuing chain is irreducible, it is recurrent.*

The proof of these results will be given in the appendix.

APPENDIX

1.9. Proof of results for the branching and queuing chains

In this section we will verify the results discussed in Section 1.8. To do so we need the following.

Theorem 6 *Let Φ be the probability generating function of a nonnegative integer-valued random variable ξ and set $\mu = E\xi$ (with $\mu = +\infty$ if ξ does not have finite expectation). If $\mu \leq 1$ and $P(\xi = 1) < 1$, the equation*

$$\Phi(t) = t \tag{71}$$

has no roots in $[0, 1)$. If $\mu > 1$, then (71) has a unique root ρ_0 in $[0, 1)$.

Graphs of $\Phi(t)$, $0 \le t \le 1$, in three typical cases corresponding to $\mu < 1$, $\mu = 1$, and $\mu > 1$ are shown in Figure 2. The fact that μ is the left-hand derivative of $\Phi(t)$ at $t = 1$ plays a fundamental role in the proof of Theorem 6.

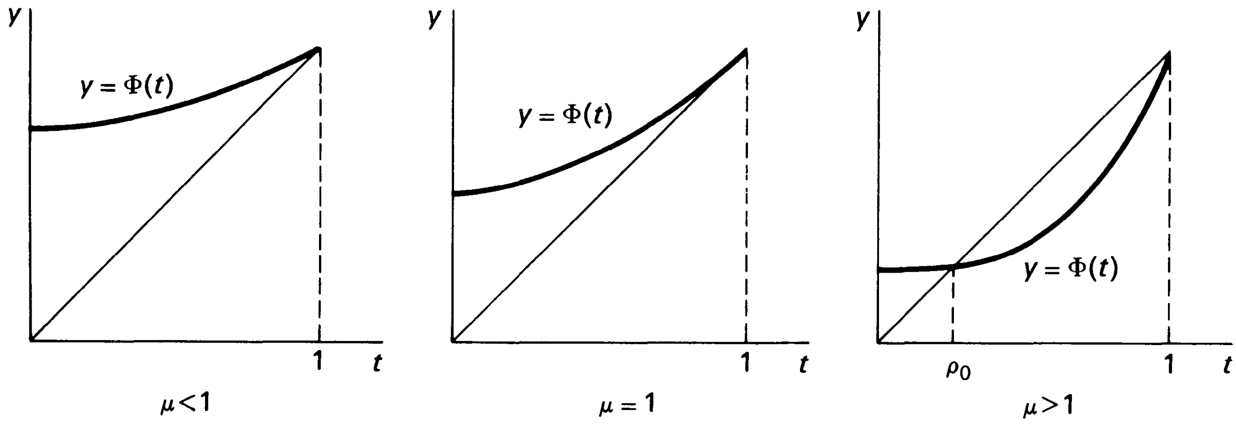


Figure 2

Proof. Let f denote the density of ξ. Then

$$\Phi(t) = f(0) + f(1)t + f(2)t^2 + \cdots$$

and

$$\Phi'(t) = f(1) + 2f(2)t + 3f(3)t^2 + \cdots.$$

Thus $\Phi(0) = f(0)$, $\Phi(1) = 1$, and

$$\lim_{t \to 1} \Phi'(t) = f(1) + 2f(2) + 3f(3) + \cdots = \mu.$$

Suppose first that $\mu < 1$. Then

$$\lim_{t \to 1} \Phi'(t) < 1.$$

Since $\Phi'(t)$ is nondecreasing in t, $0 \le t < 1$, we conclude that $\Phi'(t) < 1$ for $0 \le t < 1$. Suppose next that $\mu = 1$ and $f(1) = P(\xi = 1) < 1$. Then $f(n) > 0$ for some $n \ge 2$ (otherwise $f(0) > 0$, which implies that $\mu < 1$, a contradiction). Therefore $\Phi'(t)$ is strictly increasing in t, $0 \le t < 1$. Since

$$\lim_{t \to 1} \Phi'(t) = 1,$$

we again conclude that $\Phi'(t) < 1$ for $0 \le t < 1$.

Suppose now that $\mu \le 1$ and $P(\xi = 1) < 1$. We have shown that $\Phi'(t) < 1$ for $0 \le t < 1$. Thus

$$\frac{d}{dt}(\Phi(t) - t) < 0, \qquad 0 \le t < 1,$$

and hence $\Phi(t) - t$ is strictly decreasing on $[0, 1]$. Since $\Phi(1) - 1 = 0$, we see that $\Phi(t) - t > 0, 0 \leq t < 1$, and hence that (71) has no roots on $[0, 1)$. This proves the first part of the theorem.

Suppose next that $\mu > 1$. Then

$$\lim_{t \to 1} \Phi'(t) > 1,$$

so by the continuity of Φ' there is a number t_0 such that $0 < t_0 < 1$ and $\Phi'(t) > 1$ for $t_0 < t < 1$. It follows from the mean value theorem that

$$\frac{\Phi(1) - \Phi(t_0)}{1 - t_0} > 1.$$

Since $\Phi(1) = 1$, we conclude that $\Phi(t_0) - t_0 < 0$. Now $\Phi(t) - t$ is continuous in t and nonnegative at $t = 0$, so by the intermediate value theorem it must have a zero ρ_0 on $[0, t_0)$. Thus (71) has a root ρ_0 in $[0, 1)$. We will complete the proof of the theorem by showing that there is only one such root.

Suppose that $0 \leq \rho_0 < \rho_1 < 1$, $\Phi(\rho_0) = \rho_0$, and $\Phi(\rho_1) = \rho_1$. Then the function $\Phi(t) - t$ vanishes at ρ_0, ρ_1, and 1; hence by Rolle's theorem its first derivative has at least two roots in $(0, 1)$. By another application of Rolle's theorem its second derivative $\Phi''(t)$ has at least one root in $(0, 1)$. But if $\mu > 1$, then at least one of the numbers $f(2), f(3), \ldots$ is strictly positive, and hence

$$\Phi''(t) = 2f(2) + 3 \cdot 2f(3)t + \cdots$$

has no roots in $(0, 1)$. This contradiction shows that $\Phi(t) = t$ has a unique root in $[0, 1)$. ∎

1.9.1. Branching chain. Using Theorem 6 we see that the results for $\mu \leq 1$ follow as indicated in Section 1.8.1.

Suppose $\mu > 1$. It follows from Theorem 6 that ρ equals ρ_0 or 1, where ρ_0 is the unique root of the equation $\Phi(t) = t$ in $[0, 1)$. We will show that ρ always equals ρ_0.

First we observe that since the initial particles act independently in giving rise to their offspring, the probability $P_y(T_0 \leq n)$ that the descendants of each of the $y \geq 1$ particles become extinct by time n is given by

$$P_y(T_0 \leq n) = (P_1(T_0 \leq n))^y.$$

Consequently for $n \geq 0$ by Exercise 9(a)

$$\begin{aligned} P_1(T_0 \leq n + 1) &= P(1, 0) + \sum_{y=1}^{\infty} P(1, y)P_y(T_0 \leq n) \\ &= P(1, 0) + \sum_{y=1}^{\infty} P(1, y)(P_1(T_0 \leq n))^y \\ &= f(0) + \sum_{y=1}^{\infty} f(y)(P_1(T_0 \leq n))^y, \end{aligned}$$

and hence

(72) $$P_1(T_0 \leq n + 1) = \Phi(P_1(T_0 \leq n)), \qquad n \geq 0.$$

We will use (72) to prove by induction that

(73) $$P_1(T_0 \leq n) \leq \rho_0, \qquad n \geq 0.$$

Now

$$P_1(T_0 \leq 0) = 0 \leq \rho_0,$$

so that (73) is true for $n = 0$. Suppose that (73) holds for a given value of n. Since $\Phi(t)$ is increasing in t, we conclude from (72) that

$$P_1(T_0 \leq n + 1) = \Phi(P_1(T_0 \leq n)) \leq \Phi(\rho_0) = \rho_0,$$

and thus (73) holds for the next value of n. By induction (73) is true for all $n \geq 0$.

By letting $n \to \infty$ in (73) we see that

$$\rho = P_1(T_0 < \infty) = \lim_{n \to \infty} P_1(T_0 \leq n) \leq \rho_0.$$

Since ρ is one of the two numbers ρ_0 or 1, it must be the number ρ_0.

1.9.2. Queuing chain. We will now verify the results of Section 1.8.2. Let ξ_n denote the number of customers arriving during the nth time period. Then $\xi_1, \xi_2, \ldots$ are independent random variables having common density f, mean μ, and probability generating function Φ.

It follows from Exercise 9(b) and the identity $P(0, z) \equiv P(1, z)$, valid for a queuing chain, that $\rho_{00} = \rho_{10}$. We will show that the number $\rho = \rho_{00} = \rho_{10}$ satisfies the equation

(74) $$\Phi(\rho) = \rho.$$

If 0 is a recurrent state, $\rho = 1$ and (74) follows immediately from the fact that $\Phi(1) = 1$. To verify (74) in general, we observe first that by Exercise 9(b)

$$\rho_{00} = P(0, 0) + \sum_{y=1}^{\infty} P(0, y)\rho_{y0},$$

i.e., that

(75) $$\rho = f(0) + \sum_{y=1}^{\infty} f(y)\rho_{y0}.$$

In order to compute ρ_{y0}, $y = 1, 2, \ldots$, we consider a queuing chain starting at the positive integer y. For $n = 1, 2, \ldots$, the event $\{T_{y-1} = n\}$ occurs if and only if

$$\begin{aligned} n &= \min(m > 0: y + (\xi_1 - 1) + \cdots + (\xi_m - 1) = y - 1) \\ &= \min(m > 0: \xi_1 + \cdots + \xi_m = m - 1), \end{aligned}$$

that is, if and only if n is the smallest positive integer m such that the number of new customers entering the queue by time m is one less than the number served by time m. Thus $P_y(T_{y-1} = n)$ is independent of y, and consequently $\rho_{y,y-1} = P_y(T_{y-1} < \infty)$ is independent of y for $y = 1, 2, \ldots$. Since $\rho_{10} = \rho$, we see that

$$\rho_{y,y-1} = \rho_{y-1,y-2} = \cdots = \rho_{10} = \rho.$$

Now the queuing chain can go at most one step to the left at a time, so in order to go from state $y > 0$ to state 0 it must pass through all the intervening states $y - 1, \ldots, 1$. By applying the Markov property we can conclude (see Exercise 39) that

$$\rho_{y0} = \rho_{y,y-1}\rho_{y-1,y-2}\cdots\rho_{10} = \rho^y. \tag{76}$$

It follows from (75) and (76) that

$$\rho = f(0) + \sum_{y=1}^{\infty} f(y)\rho^y = \Phi(\rho),$$

so that (74) holds.

Using (74) and Theorem 6 it is easy to see that if $\mu \leq 1$ and the queuing chain is irreducible, then the chain is recurrent. For ρ satisfies (74) and by Theorem 6 this equation has no roots in $[0, 1)$ (observe that $P(\xi_1 = 1) < 1$ if the queuing chain is irreducible). We conclude that $\rho = 1$. Since $\rho_{00} = \rho$, state 0 is recurrent, and thus since the chain is irreducible, all states are recurrent.

Suppose now that $\mu > 1$. Again ρ satisfies (74) which, by Theorem 6, has a unique root ρ_0 in $[0, 1)$. Thus ρ equals either ρ_0 or 1. We will prove that $\rho = \rho_0$.

To this end we first observe that by Exercise 9(a)

$$P_1(T_0 \leq n + 1) = P(1, 0) + \sum_{y=1}^{\infty} P(1, y)P_y(T_0 \leq n),$$

which can be rewritten as

$$P_1(T_0 \leq n + 1) = f(0) + \sum_{y=1}^{\infty} f(y)P_y(T_0 \leq n). \tag{77}$$

We claim next that

$$P_y(T_0 \leq n) \leq (P_1(T_0 \leq n))^y, \qquad y \geq 1 \text{ and } n \geq 0. \tag{78}$$

To verify (78) observe that if a queuing chain starting at y reaches 0 in n or fewer steps, it must reach $y - 1$ in n or fewer steps, go from $y - 1$ to $y - 2$ in n or fewer steps, etc. By applying the Markov property we can conclude (see Exercise 39) that

$$P_y(T_0 \leq n) \leq P_y(T_{y-1} \leq n)P_{y-1}(T_{y-2} \leq n)\cdots P_1(T_0 \leq n). \tag{79}$$

Since

$$P_z(T_{z-1} \le n) = P_1(T_0 \le n), \qquad 1 \le z \le y,$$

(78) is valid.

It follows from (77) and (78) that

$$P_1(T_0 \le n + 1) \le f(0) + \sum_{y=1}^{\infty} f(y)(P_1(T_0 \le n))^y,$$

i.e., that

$$(80) \qquad P_1(T_0 \le n + 1) \le \Phi(P_1(T_0 \le n)), \qquad n \ge 0.$$

This in turn implies that

$$(81) \qquad P_1(T_0 \le n) \le \rho_0, \qquad n \ge 0,$$

by a proof that is almost identical to the proof that (72) implies (73) (the slight changes needed are left as an exercise for the reader). Just as in the proof of the corresponding result for the branching chain, we see by letting $n \to \infty$ in (81) that $\rho \le \rho_0$ and hence that $\rho = \rho_0$.

We have shown that if $\mu > 1$, then $\rho_{00} = \rho < 1$, and hence 0 is a transient state. It follows that if $\mu > 1$ and the chain is irreducible, then all states are transient. If $\mu > 1$ and the queuing chain is not irreducible, then case (d) of Exercise 38 holds (why?), and it is left to the reader to show that again all states are transient.

Exercises

1 Let X_n, $n \ge 0$, be the two-state Markov chain. Find
(a) $P(X_1 = 0 \mid X_0 = 0 \text{ and } X_2 = 0)$,
(b) $P(X_1 \ne X_2)$.

2 Suppose we have two boxes and $2d$ balls, of which d are black and d are red. Initially, d of the balls are placed in box 1, and the remainder of the balls are placed in box 2. At each trial a ball is chosen at random from each of the boxes, and the two balls are put back in the opposite boxes. Let X_0 denote the number of black balls initially in box 1 and, for $n \ge 1$, let X_n denote the number of black balls in box 1 after the nth trial. Find the transition function of the Markov chain X_n, $n \ge 0$.

3 Let the queuing chain be modified by supposing that if there are one or more customers waiting to be served at the start of a period, there is probability p that one customer will be served during that period and probability $1 - p$ that no customers will be served during that period. Find the transition function for this modified queuing chain.

4 Consider a probability space $(\Omega, \mathscr{A}, P)$ and assume that the various sets mentioned below are all in $\mathscr{A}$.

(a) Show that if D_i are disjoint and $P(C \mid D_i) = p$ independently of i, then $P(C \mid \bigcup_i D_i) = p$.

(b) Show that if C_i are disjoint, then $P(\bigcup_i C_i \mid D) = \sum_i P(C_i \mid D)$.

(c) Show that if E_i are disjoint and $\bigcup_i E_i = \Omega$, then

$$P(C \mid D) = \sum_i P(E_i \mid D)P(C \mid E_i \cap D).$$

(d) Show that if C_i are disjoint and $P(A \mid C_i) = P(B \mid C_i)$ for all i, then $P(A \mid \bigcup_i C_i) = P(B \mid \bigcup_i C_i)$.

5 Let X_n, $n \geq 0$, be the two-state Markov chain.

(a) Find $P_0(T_0 = n)$.

(b) Find $P_0(T_1 = n)$.

6 Let X_n, $n \geq 0$, be the Ehrenfest chain and suppose that X_0 has a binomial distribution with parameters d and $1/2$, i.e.,

$$P(X_0 = x) = \frac{\binom{d}{x}}{2^d}, \qquad x = 0, \ldots, d.$$

Find the distribution of X_1.

7 Let X_n, $n \geq 0$, be a Markov chain. Show that

$$P(X_0 = x_0 \mid X_1 = x_1, \ldots, X_n = x_n) = P(X_0 = x_0 \mid X_1 = x_1).$$

8 Let x and y be distinct states of a Markov chain having $d < \infty$ states and suppose that x leads to y. Let n_0 be the smallest positive integer such that $P^{n_0}(x, y) > 0$ and let $x_1, \ldots, x_{n_0 - 1}$ be states such that

$$P(x, x_1)P(x_1, x_2) \cdots P(x_{n_0-2}, x_{n_0-1})P(x_{n_0-1}, y) > 0.$$

(a) Show that $x, x_1, \ldots, x_{n_0-1}, y$ are distinct states.

(b) Use (a) to show that $n_0 \leq d - 1$.

(c) Conclude that $P_x(T_y \leq d - 1) > 0$.

9 Use (29) to verify the following identities:

(a) $$P_x(T_y \leq n + 1) = P(x, y) + \sum_{z \neq y} P(x, z)P_z(T_y \leq n), \qquad n \geq 0;$$

(b) $$\rho_{xy} = P(x, y) + \sum_{z \neq y} P(x, z)\rho_{zy}.$$

10 Consider the Ehrenfest chain with $d = 3$.

(a) Find $P_x(T_0 = n)$ for $x \in \mathscr{S}$ and $1 \leq n \leq 3$.

(b) Find P, P^2, and P^3.

(c) Let π_0 be the uniform distribution $\pi_0 = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$. Find π_1, π_2, and π_3.

11 Consider the genetics chain from Example 7 with $d = 3$.
(a) Find the transition matrices P and P^2.
(b) If $\pi_0 = (0, \frac{1}{2}, \frac{1}{2}, 0)$, find π_1 and π_2.
(c) Find $P_x(T_{\{0,3\}} = n)$, $x \in \mathscr{S}$, for $n = 1$ and $n = 2$.

12 Consider the Markov chain having state space $\{0, 1, 2\}$ and transition matrix

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{bmatrix} 0 & 1 & 0 \\ 1-p & 0 & p \\ 0 & 1 & 0 \end{bmatrix} \end{array}.$$

(a) Find P^2.
(b) Show that $P^4 = P^2$.
(c) Find P^n, $n \geq 1$.

13 Let X_n, $n \geq 0$, be a Markov chain whose state space $\mathscr{S}$ is a subset of $\{0, 1, 2, \ldots\}$ and whose transition function P is such that

$$\sum_y yP(x, y) = Ax + B, \qquad x \in \mathscr{S},$$

for some constants A and B.
(a) Show that $EX_{n+1} = AEX_n + B$.
(b) Show that if $A \neq 1$, then

$$EX_n = \frac{B}{1 - A} + A^n \left(EX_0 - \frac{B}{1 - A}\right).$$

14 Let X_n, $n \geq 0$, be the Ehrenfest chain on $\{0, 1, \ldots, d\}$. Show that the assumption of Exercise 13 holds and use that exercise to compute $E_x(X_n)$.

15 Let y be a transient state. Use (36) to show that for all x

$$\sum_{n=0}^{\infty} P^n(x, y) \leq \sum_{n=0}^{\infty} P^n(y, y).$$

16 Show that $\rho_{xy} > 0$ if and only if $P^n(x, y) > 0$ for some positive integer n.

17 Show that if x leads to y and y leads to z, then x leads to z.

18 Consider a Markov chain on the nonnegative integers such that, starting from x, the chain goes to state $x + 1$ with probability p, $0 < p < 1$, and goes to state 0 with probability $1 - p$.
(a) Show that this chain is irreducible.
(b) Find $P_0(T_0 = n)$, $n \geq 1$.
(c) Show that the chain is recurrent.

19 Consider a Markov chain having state space $\{0, 1, \ldots, 6\}$ and transition matrix

$$\begin{array}{c c}
 & \begin{array}{ccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\
\begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} & \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{8} & \frac{1}{4} & \frac{1}{8} & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}
\end{array}.$$

(a) Determine which states are transient and which states are recurrent.
(b) Find ρ_{0y}, $y = 0, \ldots, 6$.

20 Consider the Markov chain on $\{0, 1, \ldots, 5\}$ having transition matrix

$$\begin{array}{c c}
 & \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & 5 \end{array} \\
\begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} & \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{8} & 0 & \frac{7}{8} & 0 \\ \frac{1}{4} & \frac{1}{4} & 0 & 0 & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & \frac{3}{4} & 0 & \frac{1}{4} & 0 \\ 0 & \frac{1}{5} & 0 & \frac{1}{5} & \frac{1}{5} & \frac{2}{5} \end{bmatrix}
\end{array}.$$

(a) Determine which states are transient and which are recurrent.
(b) Find $\rho_{\{0,1\}}(x)$, $x = 0, \ldots, 5$.

21 Consider a Markov chain on $\{0, 1, \ldots, d\}$ satisfying (51) and having no absorbing states other than 0 and d. Show that the states $1, \ldots, d - 1$ each lead to 0, and hence that each is a transient state.

22 Show that the genetics chain introduced in Example 7 satisfies Equation (51).

23 A certain Markov chain that arises in genetics has states $0, 1, \ldots, 2d$ and transition function

$$P(x, y) = \binom{2d}{y} \left(\frac{x}{2d}\right)^y \left(1 - \frac{x}{2d}\right)^{2d-y}.$$

Find $\rho_{\{0\}}(x)$, $0 < x < 2d$.

24 Consider a gambler's ruin chain on $\{0, 1, \ldots, d\}$. Find

$$P_x(T_0 < T_d), \qquad 0 < x < d.$$

25 A gambler playing roulette makes a series of one dollar bets. He has respective probabilities 9/19 and 10/19 of winning and losing each bet. The gambler decides to quit playing as soon as he either is one dollar ahead or has lost his initial capital of \$1000.

(a) Find the probability that when he quits playing he will have lost \$1000.
(b) Find his expected loss.

26 Consider a birth and death chain on the nonnegative integers such that $p_x > 0$ and $q_x > 0$ for $x \geq 1$.
(a) Show that if $\sum_{y=0}^{\infty} \gamma_y = \infty$, then $\rho_{x0} = 1$, $x \geq 1$.
(b) Show that if $\sum_{y=0}^{\infty} \gamma_y < \infty$, then

$$\rho_{x0} = \frac{\sum_{y=x}^{\infty} \gamma_y}{\sum_{y=0}^{\infty} \gamma_y}, \qquad x \geq 1.$$

27 Consider a gambler's ruin chain on $\{0, 1, 2, \ldots\}$.
(a) Show that if $q \geq p$, then $\rho_{x0} = 1$, $x \geq 1$.
(b) Show that if $q < p$, then $\rho_{x0} = (q/p)^x$, $x \geq 1$.
Hint: Use Exercise 26.

28 Consider an irreducible birth and death chain on the nonnegative integers. Show that if $p_x \leq q_x$ for $x \geq 1$, the chain is recurrent.

29 Consider an irreducible birth and death chain on the nonnegative integers such that

$$\frac{q_x}{p_x} = \left(\frac{x}{x+1}\right)^2, \qquad x \geq 1.$$

(a) Show that this chain is transient.
(b) Find ρ_{x0}, $x \geq 1$. *Hint:* Use Exercise 26 and the formula $\sum_{y=1}^{\infty} 1/y^2 = \pi^2/6$.

30 Consider the birth and death chain in Example 13.
(a) Compute $P_x(T_a < T_b)$ for $a < x < b$.
(b) Compute ρ_{x0}, $x > 0$.

31 Consider a branching chain such that $f(1) < 1$. Show that every state other than 0 is transient.

32 Consider the branching chain described in Example 14. If a given man has two boys and one girl, what is the probability that his male line will continue forever?

33 Consider a branching chain with $f(0) = f(3) = 1/2$. Find the probability ρ of extinction.

34 Consider a branching chain with $f(x) = p(1 - p)^x$, $x \geq 0$, where $0 < p < 1$. Show that $\rho = 1$ if $p \geq 1/2$ and that $\rho = p/(1 - p)$ if $p < 1/2$.

35 Let X_n, $n \geq 0$, be a branching chain. Show that $E_x(X_n) = x\mu^n$.
Hint: See Exercise 13.

36 Let X_n, $n \geq 0$, be a branching chain and suppose that the associated random variable ξ has finite variance σ^2.
(a) Show that

$$E[X_{n+1}^2 \mid X_n = x] = x\sigma^2 + x^2\mu^2.$$

(b) Use Exercise 35 to show that

$$E_x(X_{n+1}^2) = x\mu^n\sigma^2 + \mu^2 E_x(X_n^2).$$

Hint: Use the formula $EY = \sum_x P(X = x)E[Y \mid X = x]$.

(c) Show that

$$E_x(X_n^2) = x\sigma^2(\mu^{n-1} + \cdots + \mu^{2(n-1)}) + x^2\mu^{2n}, \qquad n \geq 1.$$

(d) Show that if there are x particles initially, then for $n \geq 1$

$$\operatorname{Var} X_n = \begin{cases} x\sigma^2\mu^{n-1}\left(\dfrac{1-\mu^n}{1-\mu}\right), & \mu \neq 1, \\ nx\sigma^2, & \mu = 1. \end{cases}$$

37 Consider the queuing chain.

(a) Show that if either $f(0) = 0$ or $f(0) + f(1) = 1$, the chain is not irreducible.

(b) Show that if $f(0) > 0$ and $f(0) + f(1) < 1$, the chain is irreducible. *Hint:* First verify that (i) $\rho_{xy} > 0$ for $0 \leq y < x$; and (ii) if $x_0 \geq 2$ and $f(x_0) > 0$, then $\rho_{0,x_0+n(x_0-1)} > 0$ for $n \geq 0$.

38 Determine which states of the queuing chain are absorbing, which are recurrent, and which are transient, when the chain is not irreducible. Consider the following four cases separately (see Exercise 37):

(a) $f(1) = 1$;
(b) $f(0) > 0, f(1) > 0$, and $f(0) + f(1) = 1$;
(c) $f(0) = 1$;
(d) $f(0) = 0$ and $f(1) < 1$.

39 Consider the queuing chain.

(a) Show that for $y \geq 2$ and m a positive integer

$$P_y(T_0 = m) = \sum_{k=1}^{m-1} P_y(T_{y-1} = k)P_{y-1}(T_0 = m - k).$$

(b) By summing the equation in (a) on $m = 1, 2, \ldots$, show that

$$\rho_{y0} = \rho_{y,y-1}\,\rho_{y-1,0} \qquad y \geq 2.$$

(c) Why does Equation (76) follow from (b)?

(d) By summing the equation in (a) on $m = 1, 2, \ldots, n$, show that

$$P_y(T_0 \leq n) \leq P_y(T_{y-1} \leq n)P_{y-1}(T_0 \leq n), \qquad y \geq 2.$$

(e) Why does Equation (79) follow from (d)?

40 Verify that (81) follows from (80) by induction.

2 Stationary Distributions of a Markov Chain

Let X_n, $n \geq 0$, be a Markov chain having state space $\mathscr{S}$ and transition function P. If $\pi(x)$, $x \in \mathscr{S}$, are nonnegative numbers summing to one, and if

$$\sum_x \pi(x)P(x, y) = \pi(y), \qquad y \in \mathscr{S}, \tag{1}$$

then π is called a *stationary distribution*. Suppose that a stationary distribution π exists and that

$$\lim_{n \to \infty} P^n(x, y) = \pi(y), \qquad y \in \mathscr{S}. \tag{2}$$

Then, as we will soon see, regardless of the initial distribution of the chain, the distribution of X_n approaches π as $n \to \infty$. In such cases, π is sometimes called the *steady state distribution.*

In this chapter we will determine which Markov chains have stationary distributions, when there is such a unique distribution, and when (2) holds.

2.1. Elementary properties of stationary distributions

Let π be a stationary distribution. Then

$$\begin{aligned}\sum_x \pi(x)P^2(x, y) &= \sum_x \pi(x) \sum_z P(x, z)P(z, y) \\ &= \sum_z \left(\sum_x \pi(x)P(x, z)\right) P(z, y) \\ &= \sum_z \pi(z)P(z, y) = \pi(y).\end{aligned}$$

Similarly by induction based on the formula

$$P^{n+1}(x, y) = \sum_z P^n(x, z)P(z, y),$$

we conclude that for all n

$$\sum_x \pi(x)P^n(x, y) = \pi(y), \qquad y \in \mathscr{S}. \tag{3}$$

If X_0 has the stationary distribution π for its initial distribution, then (3) implies that for all n

$$P(X_n = y) = \pi(y), \qquad y \in \mathcal{S}, \tag{4}$$

and hence that the distribution of X_n is independent of n. Suppose conversely that the distribution of X_n is independent of n. Then the initial distribution π_0 is such that

$$\pi_0(y) = P(X_0 = y) = P(X_1 = y) = \sum_x \pi_0(x)P(x, y).$$

Consequently π_0 is a stationary distribution. In summary, the distribution of X_n is independent of n if and only if the initial distribution is a stationary distribution.

Suppose now that π is a stationary distribution and that (2) holds. Let π_0 be the initial distribution. Then

$$P(X_n = y) = \sum_x \pi_0(x)P^n(x, y), \qquad y \in \mathcal{S}. \tag{5}$$

By using (2) and the bounded convergence theorem stated in Section 2.5, we can let $n \to \infty$ in (5), obtaining

$$\lim_{n \to \infty} P(X_n = y) = \sum_x \pi_0(x)\pi(y).$$

Since $\sum_x \pi_0(x) = 1$, we conclude that

$$\lim_{n \to \infty} P(X_n = y) = \pi(y), \qquad y \in \mathcal{S}. \tag{6}$$

Formula (6) states that, regardless of the initial distribution, for large values of n the distribution of X_n is approximately equal to the stationary distribution π. It implies that π is the unique stationary distribution. For if there were some other stationary distribution we could use it for the initial distribution π_0. From (4) and (6) we would conclude that $\pi_0(y) = \pi(y)$, $y \in \mathcal{S}$.

Consider a system described by a Markov chain having transition function P and unique stationary distribution π. Suppose we start observing the system after it has been going on for some time, say n_0 units of time for some large positive integer n_0. In effect, we observe Y_n, $n \geq 0$, where

$$Y_n = X_{n+n_0}, \qquad n \geq 0.$$

The random variables Y_n, $n \geq 0$, also form a Markov chain with transition function P. In order to determine unique probabilities for events defined in terms of the Y_n chain, we need to know its initial distribution, which is the same as the distribution of X_{n_0}. In most practical applications it is very hard to determine this distribution exactly. We may have no choice but to assume that Y_n, $n \geq 0$, has the stationary distribution π for its initial distribution. This is a reasonable assumption if (2) holds and n_0 is large.

2.2. Examples

In this section we will consider some examples in which we can show directly that a unique stationary distribution exists and find simple formulas for it.

In Section 1.1 we discussed the two-state Markov chain on $\mathscr{S} = \{0, 1\}$ having transition matrix

$$\begin{array}{c} \\ 0 \\ 1 \end{array}\begin{array}{c} \begin{array}{cc} 0 & \quad 1 \end{array} \\ \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix} \end{array}.$$

We saw that if $p + q > 0$, the chain has a unique stationary distribution π, determined by

$$\pi(0) = \frac{q}{p+q} \quad \text{and} \quad \pi(1) = \frac{p}{p+q}.$$

We also saw that if $0 < p + q < 2$, then (2) holds.

For Markov chains having a finite number of states, stationary distributions can be found by solving a finite system of linear equations.

Example 1. Consider a Markov chain having state space $\mathscr{S} = \{0, 1, 2\}$ and transition matrix

$$\begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array}\begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{6} & \frac{1}{3} & \frac{1}{2} \end{bmatrix} \end{array}.$$

Show that this chain has a unique stationary distribution π and find π.

Formula (1) in this case gives us the three equations

$$\frac{\pi(0)}{3} + \frac{\pi(1)}{4} + \frac{\pi(2)}{6} = \pi(0),$$

$$\frac{\pi(0)}{3} + \frac{\pi(1)}{2} + \frac{\pi(2)}{3} = \pi(1),$$

$$\frac{\pi(0)}{3} + \frac{\pi(1)}{4} + \frac{\pi(2)}{2} = \pi(2).$$

$\sum_x \pi(x) = 1$ gives us the fourth equation

$$\pi(0) + \pi(1) + \pi(2) = 1.$$

By subtracting twice the first equation from the second equation, we eliminate the term involving $\pi(2)$ and find that $\pi(1) = 5\pi(0)/3$. We conclude from the first equation that $\pi(2) = 3\pi(0)/2$. From the fourth equation we now see that

$$\pi(0)(1 + \tfrac{5}{3} + \tfrac{3}{2}) = 1,$$

and hence that

$$\pi(0) = \tfrac{6}{25}.$$

Thus

$$\pi(1) = \tfrac{5}{3} \cdot \tfrac{6}{25} = \tfrac{2}{5}$$

and

$$\pi(2) = \tfrac{3}{2} \cdot \tfrac{6}{25} = \tfrac{9}{25}.$$

It is readily seen that these numbers satisfy all four equations. Since they are nonnegative, the unique stationary distribution is given by

$$\pi(0) = \tfrac{6}{25}, \qquad \pi(1) = \tfrac{2}{5}, \qquad \text{and} \qquad \pi(2) = \tfrac{9}{25}.$$

Though it is not easy to see directly, (2) holds for this chain (see Section 2.7).

2.2.1. Birth and death chain. Consider a birth and death chain on $\{0, 1, \ldots, d\}$ or on the nonnegative integers. In the latter case we set $d = \infty$. We assume without further mention that the chain is irreducible, i.e., that

$$p_x > 0 \qquad \text{for} \qquad 0 \le x < d$$

and

$$q_x > 0 \qquad \text{for} \qquad 0 < x \le d$$

if d is finite, and that

$$p_x > 0 \qquad \text{for} \qquad 0 \le x < \infty$$

and

$$q_x > 0 \qquad \text{for} \qquad 0 < x < \infty$$

if d is infinite.

Suppose d is infinite. The system of equations

$$\sum_x \pi(x)P(x, y) = \pi(y), \qquad y \in \mathscr{S},$$

becomes

$$\pi(0)r_0 + \pi(1)q_1 = \pi(0),$$
$$\pi(y-1)p_{y-1} + \pi(y)r_y + \pi(y+1)q_{y+1} = \pi(y), \qquad y \geq 1.$$

Since

$$p_y + q_y + r_y = 1,$$

these equations reduce to

(7)
$$q_1\pi(1) - p_0\pi(0) = 0,$$
$$q_{y+1}\pi(y+1) - p_y\pi(y) = q_y\pi(y) - p_{y-1}\pi(y-1), \qquad y \geq 1.$$

It follows easily from (7) and induction that

$$q_{y+1}\pi(y+1) - p_y\pi(y) = 0, \qquad y \geq 0,$$

and hence that

$$\pi(y+1) = \frac{p_y}{q_{y+1}}\pi(y), \qquad y \geq 0.$$

Consequently,

(8)
$$\pi(x) = \frac{p_0 \cdots p_{x-1}}{q_1 \cdots q_x}\pi(0), \qquad x \geq 1.$$

Set

(9)
$$\pi_x = \begin{cases} 1, & x = 0, \\ \dfrac{p_0 \cdots p_{x-1}}{q_1 \cdots q_x}, & x \geq 1. \end{cases}$$

Then (8) can be written as

(10)
$$\pi(x) = \pi_x\pi(0), \qquad x \geq 0.$$

Conversely, (1) follows from (10).

Suppose now that $\sum_x \pi_x < \infty$ or, equivalently, that

(11)
$$\sum_{x=1}^{\infty} \frac{p_0 \cdots p_{x-1}}{q_1 \cdots q_x} < \infty.$$

We conclude from (10) that the birth and death chain has a unique stationary distribution, given by

(12)
$$\pi(x) = \frac{\pi_x}{\sum_{y=0}^{\infty} \pi_y}, \qquad x \geq 0.$$

Suppose instead that (11) fails to hold, i.e., that

(13)
$$\sum_{x=0}^{\infty} \pi_x = \infty.$$

We conclude from (10) and (13) that any solution to (1) is either identically zero or has infinite sum, and hence that there is no stationary distribution.

In summary, we see that the chain has a stationary distribution if and only if (11) holds, and that the stationary distribution, when it exists, is given by (9) and (12).

Suppose now that $d < \infty$. By essentially the same arguments used to obtain (12), we conclude that the unique stationary distribution is given by

$$(14) \qquad \pi(x) = \frac{\pi_x}{\sum_{y=0}^{d} \pi_y}, \qquad 0 \le x \le d,$$

where π_x, $0 \le x \le d$, is given by (9).

Example 2. Consider the Ehrenfest chain introduced in Section 1.3 and suppose that $d = 3$. Find the stationary distribution.

The transition matrix of the chain is

$$\begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} & 0 \\ 0 & \frac{2}{3} & 0 & \frac{1}{3} \\ 0 & 0 & 1 & 0 \end{bmatrix} \end{array}.$$

This is an irreducible birth and death chain in which $\pi_0 = 1$,

$$\pi_1 = \frac{1}{\frac{1}{3}} = 3, \qquad \pi_2 = \frac{1 \cdot \frac{2}{3}}{\frac{1}{3} \cdot \frac{2}{3}} = 3,$$

and

$$\pi_3 = \frac{1 \cdot \frac{2}{3} \cdot \frac{1}{3}}{\frac{1}{3} \cdot \frac{2}{3} \cdot 1} = 1.$$

Thus the unique stationary distribution is given by

$$\pi(0) = \tfrac{1}{8}, \qquad \pi(1) = \tfrac{3}{8}, \qquad \pi(2) = \tfrac{3}{8}, \qquad \text{and} \qquad \pi(3) = \tfrac{1}{8}.$$

Formula (2) does not hold for the chain in Example 2 since $P^n(x, x) = 0$ for odd values of n. We can modify the Ehrenfest chain slightly and avoid such "periodic" behavior.

Example 3. Modified Ehrenfest chain. Suppose we have two boxes labeled 1 and 2 and d balls labeled $1, 2, \ldots, d$. Initially some of the balls are in box 1 and the remainder are in box 2. An integer is selected at random from $1, 2, \ldots, d$, and the ball labeled by that integer is removed from its box. We now select at random one of the two boxes and put the removed ball into this box. The procedure is repeated indefinitely, the selections being made independently. Let X_n denote the number of balls in box 1 after the nth trial. Then X_n, $n \geq 0$, is a Markov chain on $\mathscr{S} = \{0, 1, \ldots, d\}$. Find the stationary distribution of the chain for $d = 3$.

The transition matrix of this chain, for $d = 3$, is

$$\begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array}\begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3} & 0 \\ 0 & \frac{1}{3} & \frac{1}{2} & \frac{1}{6} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \end{array}.$$

To see why P is given as indicated, we will compute $P(1, y)$, $0 \leq y \leq 3$. We start with one ball in box 1 and two balls in box 2. Thus $P(1, 0)$ is the probability that the ball selected is from box 1 and the box selected is box 2. Thus

$$P(1, 0) = \tfrac{1}{3} \cdot \tfrac{1}{2} = \tfrac{1}{6}.$$

Secondly, $P(1, 2)$ is the probability that the ball selected is from box 2 and the box selected is box 1. Thus

$$P(1, 2) = \tfrac{2}{3} \cdot \tfrac{1}{2} = \tfrac{1}{3}.$$

Clearly $P(1, 3) = 0$, since at most one ball is transferred at a time. Finally, $P(1, 1)$ can be obtained by subtracting $P(1, 0) + P(1, 2) + P(1, 3)$ from 1. Alternatively, $P(1, 1)$ is the probability that either the selected ball is from box 1 and the selected box is box 1 or the selected ball is from box 2 and the selected box is box 2. Thus

$$P(1, 1) = \tfrac{1}{3} \cdot \tfrac{1}{2} + \tfrac{2}{3} \cdot \tfrac{1}{2} = \tfrac{1}{2}.$$

The other probabilities are computed similarly. This Markov chain is an irreducible birth and death chain. It is easily seen that π_x, $0 \leq x \leq 3$, are the same as in the previous example and hence that the stationary distribution is again given by

$$\pi(0) = \tfrac{1}{8}, \qquad \pi(1) = \tfrac{3}{8}, \qquad \pi(2) = \tfrac{3}{8}, \qquad \text{and} \qquad \pi(3) = \tfrac{1}{8}.$$

It follows from the results in Section 2.7 that (2) holds for the chain in Example 3.

2.2.2. Particles in a box. A Markov chain that arises in several applied contexts can be described as follows. Suppose that ξ_n particles are added to a box at times $n = 1, 2, \ldots$, where ξ_n, $n \geq 1$, are independent and have a Poisson distribution with common parameter λ. Suppose that each particle in the box at time n, independently of all the other particles in the box and independently of how particles are added to the box, has probability $p < 1$ of remaining in the box at time $n + 1$ and probability $q = 1 - p$ of being removed from the box at time $n + 1$. Let X_n denote the number of particles in the box at time n. Then X_n, $n \geq 0$, is a Markov chain. We will find the stationary distribution of this chain. We will also find an explicit formula for $P^n(x, y)$ and use this formula to show directly that (2) holds.

The same Markov chain can be used to describe a telephone exchange, where ξ_n is the number of new calls starting at time n, q is the probability that a call in progress at time n terminates by time $n + 1$, and X_n is the number of calls in progress at time n.

We will now analyze this Markov chain. Let $R(X_n)$ denote the number of particles present at time n that remain in the box at time $n + 1$. Then

$$X_{n+1} = \xi_{n+1} + R(X_n).$$

Clearly

$$P(R(X_n) = z \mid X_n = x) = \binom{x}{z} p^z(1 - p)^{x-z}, \qquad 0 \leq z \leq x,$$

and

$$P(\xi_n = z) = \frac{\lambda^z e^{-\lambda}}{z!}, \qquad z \geq 0.$$

Since

$$\begin{aligned} P(X_{n+1} = y \mid X_n = x) &= \sum_{z=0}^{\min(x,y)} P(R(X_n) = z, \xi_{n+1} = y - z \mid X_n = x) \\ &= \sum_{z=0}^{\min(x,y)} P(\xi_{n+1} = y - z)P(R(X_n) = z \mid X_n = x), \end{aligned}$$

we conclude that

$$P(x, y) = \sum_{z=0}^{\min(x,y)} \frac{\lambda^{y-z}e^{-\lambda}}{(y - z)!} \binom{x}{z} p^z(1 - p)^{x-z}. \tag{15}$$

It follows from (15) or from the original description of the process that $P(x, y) > 0$ for all $x \geq 0$ and $y \geq 0$, and hence that the chain is irreducible.

Suppose X_n has a Poisson distribution with parameter t. Then $R(X_n)$ has a Poisson distribution with parameter pt. For

$$\begin{aligned} P(R(X_n) = y) &= \sum_{x=y}^{\infty} P(X_n = x, R(X_n) = y) \\ &= \sum_{x=y}^{\infty} P(X_n = x)P(R(X_n) = y \mid X_n = x) \\ &= \sum_{x=y}^{\infty} \frac{t^x e^{-t}}{x!} \binom{x}{y} p^y(1 - p)^{x-y} \end{aligned}$$

$$
\begin{aligned}
&= \sum_{x=y}^{\infty} \frac{t^x e^{-t}}{y!\,(x-y)!}\, p^y(1-p)^{x-y} \\
&= \frac{(pt)^y e^{-t}}{y!} \sum_{x=y}^{\infty} \frac{(t(1-p))^{x-y}}{(x-y)!} \\
&= \frac{(pt)^y e^{-t}}{y!} \sum_{z=0}^{\infty} \frac{(t(1-p))^z}{z!} \\
&= \frac{(pt)^y e^{-t}}{y!}\, e^{t(1-p)} \\
&= \frac{(pt)^y e^{-pt}}{y!},
\end{aligned}
$$

which shows that $R(X_n)$ has the indicated Poisson distribution.

We will now show that the stationary distribution is Poisson with parameter t for suitable t. Let X_0 have such a distribution. Then $X_1 = \xi_1 + R(X_0)$ is the sum of independent random variables having Poisson distributions with parameters λ and pt respectively. Thus X_1 has a Poisson distribution with parameter $\lambda + pt$. The distribution of X_1 will agree with that of X_0 if $t = \lambda + pt$, i.e., if

$$t = \frac{\lambda}{1-p} = \frac{\lambda}{q}.$$

We conclude that the Markov chain has a stationary distribution π which is a Poisson distribution with parameter λ/q, i.e., such that

$$(16) \qquad \pi(x) = \frac{(\lambda/q)^x\, e^{-\lambda/q}}{x!}, \qquad x \geq 0.$$

Finally we will derive a formula for $P^n(x, y)$. Suppose X_0 has a Poisson distribution with parameter t. It is left as an exercise for the reader to show that X_n has a Poisson distribution with parameter

$$tp^n + \frac{\lambda}{q}(1 - p^n).$$

Thus

$$
\begin{aligned}
\sum_{x=0}^{\infty} \frac{e^{-t}t^x}{x!} P^n(x, y) &= P(X_n = y) \\
&= \exp\left[-\left(tp^n + \frac{\lambda}{q}(1-p^n)\right)\right] \frac{\left[tp^n + \frac{\lambda}{q}(1-p^n)\right]^y}{y!},
\end{aligned}
$$

and hence

$$(17) \qquad \sum_{x=0}^{\infty} t^x \frac{P^n(x, y)}{x!} = e^{-\lambda(1-p^n)/q} e^{t(1-p^n)} \frac{\left[tp^n + \frac{\lambda}{q}(1-p^n)\right]^y}{y!}.$$

Now if

$$\sum_{x=0}^{\infty} c_x t^x = \left(\sum_{x=0}^{\infty} b_x t^x\right)\left(\sum_{x=0}^{\infty} a_x t^x\right),$$

where each power series has a positive radius of convergence, then

$$c_x = \sum_{z=0}^{x} a_z b_{x-z}.$$

If $a_z = 0$ for $z > y$, then

$$c_x = \sum_{z=0}^{\min(x,y)} a_z b_{x-z}.$$

Using this with (17) and the binomial expansion, we conclude that

$$P^n(x, y) = \frac{x!\, e^{-\lambda(1-p^n)/q}}{y!} \sum_{z=0}^{\min(x,y)} \binom{y}{z} p^{nz} \left[\frac{\lambda}{q}(1 - p^n)\right]^{y-z} \frac{(1 - p^n)^{x-z}}{(x - z)!},$$

which simplifies slightly to

$$(18) \qquad P^n(x, y) = e^{-\lambda(1-p^n)/q} \sum_{z=0}^{\min(x,y)} \binom{x}{z} p^{nz}(1 - p^n)^{x-z} \frac{\left[\frac{\lambda}{q}(1 - p^n)\right]^{y-z}}{(y - z)!}.$$

Since $0 \le p < 1$,

$$\lim_{n\to\infty} p^n = 0.$$

Thus as $n \to \infty$, the terms in the sum in (18) all approach zero except for the term corresponding to $z = 0$. We conclude that

$$(19) \qquad \lim_{n\to\infty} P^n(x, y) = \frac{e^{-\lambda/q}\left(\frac{\lambda}{q}\right)^y}{y!} = \pi(y), \qquad x, y \ge 0.$$

Thus (2) holds for this chain, and consequently the distribution π given by (16) is the unique stationary distribution of the chain.

2.3. Average number of visits to a recurrent state

Consider an irreducible birth and death chain with stationary distribution π. Suppose that $P(x, x) = r_x = 0$, $x \in \mathscr{S}$, as in the Ehrenfest chain and the gambler's ruin chain. Then at each transition the birth and death chain moves either one step to the right or one step to the left. Thus the chain can return to its starting point only after an even number of transitions. In other words, $P^n(x, x) = 0$ for odd values of n. For such a chain the formula

$$\lim_{n\to\infty} P^n(x, y) = \pi(y), \qquad y \in \mathscr{S},$$

clearly fails to hold.

There is a way to handle such situations. Let a_n, $n \geq 0$, be a sequence of numbers. If

$$\lim_{n\to\infty} a_n = L \tag{20}$$

for some finite number L, then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} a_m = L. \tag{21}$$

Formula (21) can hold, however, even if (20) fails to hold. For example, if $a_n = 0$ for n odd and $a_n = 1$ for n even, then a_n has no limit as $n \to \infty$, but

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} a_m = \frac{1}{2}.$$

In this section we will show that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} P^m(x, y)$$

exists for every pair x, y of states for an arbitrary Markov chain. In Section 2.5 we will use the existence of these limits to determine which Markov chains have stationary distributions and when there is such a unique distribution.

Recall that

$$1_y(z) = \begin{cases} 1, & z = y, \\ 0, & z \neq y, \end{cases}$$

and that

$$E_x(1_y(X_n)) = P_x(X_n = y) = P^n(x, y). \tag{22}$$

Set

$$N_n(y) = \sum_{m=1}^{n} 1_y(X_m)$$

and

$$G_n(x, y) = \sum_{m=1}^{n} P^m(x, y).$$

Then $N_n(y)$ denotes the number of visits of the Markov chain to y during times $m = 1, \ldots, n$. The expected number of such visits for a chain starting at x is given according to (22) by

$$E_x(N_n(y)) = G_n(x, y). \tag{23}$$

Let y be a transient state. Then

$$\lim_{n\to\infty} N_n(y) = N(y) < \infty \qquad \text{with probability one,}$$

and

$$\lim_{n\to\infty} G_n(x, y) = G(x, y) < \infty, \qquad x \in \mathscr{S}.$$

It follows that

$$\lim_{n\to\infty} \frac{N_n(y)}{n} = 0 \qquad \text{with probability one,} \tag{24}$$

and that

$$\lim_{n\to\infty} \frac{G_n(x, y)}{n} = 0, \qquad x \in \mathscr{S}. \tag{25}$$

Observe that $N_n(y)/n$ is the proportion of the first n units of time that the chain is in state y and that $G_n(x, y)/n$ is the expected value of this proportion for a chain starting at x.

Suppose now that y is a recurrent state. Let $m_y = E_y(T_y)$ denote the *mean return time* to y for a chain starting at y if this return time has finite expectation, and set $m_y = \infty$ otherwise. Let $1_{\{T_y < \infty\}}$ denote the random variable that is 1 if $T_y < \infty$ and 0 if $T_y = \infty$.

We will use the strong law of large numbers to prove the main result of this section, namely, Theorem 1 below.

Strong Law of Large Numbers. *Let $\xi_1, \xi_2, \ldots$ be independent identically distributed random variables. If these random variables have finite mean μ, then*

$$\lim_{n\to\infty} \frac{\xi_1 + \cdots + \xi_n}{n} = \mu \qquad \textit{with probability one.}$$

If these random variables are nonnegative and fail to have finite expectation, then this limit holds, provided that we set $\mu = +\infty$.

This important theorem is proved in advanced probability texts.

Theorem 1 *Let y be a recurrent state. Then*

$$\lim_{n\to\infty} \frac{N_n(y)}{n} = \frac{1_{\{T_y < \infty\}}}{m_y} \qquad \textit{with probability one,} \tag{26}$$

and

$$\lim_{n\to\infty} \frac{G_n(x, y)}{n} = \frac{\rho_{xy}}{m_y}, \qquad x \in \mathscr{S}. \tag{27}$$

These formulas are intuitively very reasonable. Once a chain reaches y, it returns to y "on the average every m_y units of time." Thus if $T_y < \infty$ and n is large, the proportion of the first n units of time that the chain is in state y should be about $1/m_y$. Formula (27) should follow from (26) by taking expectations.

From Corollary 1 of Chapter 1 and the above theorem, we immediately obtain the next result.

Corollary 1 *Let C be an irreducible closed set of recurrent states. Then*

$$(28) \qquad \lim_{n\to\infty} \frac{G_n(x, y)}{n} = \frac{1}{m_y}, \qquad x, y \in C,$$

and if $P(X_0 \in C) = 1$, then with probability one

$$(29) \qquad \lim_{n\to\infty} \frac{N_n(y)}{n} = \frac{1}{m_y}, \qquad y \in C.$$

If $m_y = \infty$ the right sides of (26)–(29) all equal zero, and hence (24) and (25) hold.

Proof. In order to verify Theorem 1, we need to introduce some additional random variables. Consider a Markov chain starting at a recurrent state y. With probability one it returns to y infinitely many times. For $r \geq 1$ let T_y^r denote the time of the rth visit to y, so that

$$T_y^r = \min (n \geq 1 : N_n(y) = r).$$

Set $W_y^1 = T_y^1 = T_y$ and for $r \geq 2$ let $W_y^r = T_y^r - T_y^{r-1}$ denote the waiting time between the $(r-1)$th visit to y and the rth visit to y. Clearly

$$T_y^r = W_y^1 + \cdots + W_y^r.$$

The random variables $W_y^1, W_y^2, \ldots$ are independent and identically distributed and hence they have common mean $E_y(W_y^1) = E_y(T_y) = m_y$. This result should be intuitively obvious, since every time the chain returns to y it behaves from then on just as would a chain starting out initially at y. One can give a rigorous proof of this result by using (27) of Chapter 1 to show that for $r \geq 1$

$$P(W_y^{r+1} = m_{r+1} \mid W_y^1 = m_1, \ldots, W_y^r = m_r) = P_y(W_y^1 = m_{r+1});$$

and then showing by induction that

$$P_y(W_y^1 = m_1, \ldots, W_y^r = m_r) = P_y(W_y^1 = m_1) \cdots P_y(W_y^1 = m_r).$$

The strong law of large numbers implies that

$$\lim_{k\to\infty} \frac{W_y^1 + W_y^2 + \cdots + W_y^k}{k} = m_y \qquad \text{with probability one,}$$

i.e., that

$$(30) \qquad \lim_{k\to\infty} \frac{T_y^k}{k} = m_y \qquad \text{with probability one.}$$

Set $r = N_n(y)$. By time n the chain has made exactly r visits to y. Thus the rth visit to y occurs on or before time n, and the $(r + 1)$th visit to y occurs after time n; that is,

$$T_y^{N_n(y)} \le n < T_y^{N_n(y)+1},$$

and hence

$$\frac{T_y^{N_n(y)}}{N_n(y)} \le \frac{n}{N_n(y)} \le \frac{T_y^{N_n(y)+1}}{N_n(y)},$$

or at least these results hold for n large enough so that $N_n(y) \ge 1$. Since $N_n(y) \to \infty$ with probability one as $n \to \infty$, these inequalities and (30) together imply that

$$\lim_{n\to\infty} \frac{n}{N_n(y)} = m_y \qquad \text{with probability one,}$$

or, equivalently, that (29) holds.

Let y be a recurrent state as before, but let X_0 have an arbitrary distribution. Then the chain may never reach y. If it does reach y, however, the above argument is valid; and hence, with probability one, $N_n(y)/n \to 1\{T_y < \infty\}/m_y$ as $n \to \infty$. Thus (26) is valid.

By definition $0 \le N_n(y) \le n$, and hence

$$(31) \qquad 0 \le \frac{N_n(y)}{n} \le 1.$$

A theorem from measure theory, known as the dominated convergence theorem, allows us to conclude from (26) and (31) that

$$\lim_{n\to\infty} E_x\left(\frac{N_n(y)}{n}\right) = E_x\left(\frac{1_{\{T_y<\infty\}}}{m_y}\right) = \frac{P_x(T_y < \infty)}{m_y} = \frac{\rho_{xy}}{m_y}$$

and hence from (23) that (27) holds. This completes the proof of Theorem 1. ∎

2.4. Null recurrent and positive recurrent states

A recurrent state y is called *null recurrent* if $m_y = \infty$. From Theorem 1 we see that if y is null recurrent, then

$$(32) \qquad \lim_{n\to\infty} \frac{G_n(x, y)}{n} = \lim_{n\to\infty} \frac{\sum_{m=1}^n P^m(x, y)}{n} = 0, \qquad x \in \mathscr{S}.$$

(It can be shown that if y is null recurrent, then

$$\lim_{n \to \infty} P^n(x, y) = 0, \qquad x \in \mathscr{S}, \tag{33}$$

which is a stronger result than (32). We will not prove (33), since it will not be needed later and its proof is rather difficult.)

A recurrent state y is called *positive recurrent* if $m_y < \infty$. It follows from Theorem 1 that if y is positive recurrent, then

$$\lim_{n \to \infty} \frac{G_n(y, y)}{n} = \frac{1}{m_y} > 0.$$

Thus (32) and (33) fail to hold for positive recurrent states.

Consider a Markov chain starting out in a recurrent state y. It follows from Theorem 1 that if y is null recurrent, then, with probability one, the proportion of time the chain is in state y during the first n units of time approaches zero as $n \to \infty$. On the other hand, if y is a positive recurrent state, then, with probability one, the proportion of time the chain is in state y during the first n units of time approaches the positive limit $1/m_y$ as $n \to \infty$.

The next result is closely related to Theorem 2 of Chapter 1.

Theorem 2 *Let x be a positive recurrent state and suppose that x leads to y. Then y is positive recurrent.*

Proof. It follows from Theorem 2 of Chapter 1 that y leads to x. Thus there exist positive integers n_1 and n_2 such that

$$P^{n_1}(y, x) > 0 \qquad \text{and} \qquad P^{n_2}(x, y) > 0.$$

Now

$$P^{n_1+m+n_2}(y, y) \geq P^{n_1}(y, x)P^m(x, x)P^{n_2}(x, y),$$

and by summing on $m = 1, 2, \ldots, n$ and dividing by n, we conclude that

$$\frac{G_{n_1+n+n_2}(y, y)}{n} - \frac{G_{n_1+n_2}(y, y)}{n} \geq P^{n_1}(y, x)P^{n_2}(x, y)\frac{G_n(x, x)}{n}.$$

As $n \to \infty$, the left side of this inequality converges to $1/m_y$ and the right side converges to

$$\frac{P^{n_1}(y, x)P^{n_2}(x, y)}{m_x}.$$

Hence

$$\frac{1}{m_y} \geq \frac{P^{n_1}(y, x)P^{n_2}(x, y)}{m_x} > 0,$$

and consequently $m_y < \infty$. This shows that y is positive recurrent. ∎

From this theorem and from Theorem 2 of Chapter 1 we see that if C is an irreducible closed set, then every state in C is transient, every state in C is null recurrent, or every state in C is positive recurrent. A Markov chain is called a *null recurrent chain* if all its states are null recurrent and a *positive recurrent chain* if all its states are positive recurrent. We see therefore that an irreducible Markov chain is a transient chain, a null recurrent chain, or a positive recurrent chain.

If C is a finite closed set of states, then C has at least one positive recurrent state. For

$$\sum_{y \in C} P^m(x, y) = 1, \qquad x \in C,$$

and by summing on $m = 1, \ldots, n$ and dividing by n we find that

$$\sum_{y \in C} \frac{G_n(x, y)}{n} = 1, \qquad x \in C.$$

If C is finite and each state in C is transient or null recurrent, then (25) holds and hence

$$\begin{aligned} 1 &= \lim_{n \to \infty} \sum_{y \in C} \frac{G_n(x, y)}{n} \\ &= \sum_{y \in C} \lim_{n \to \infty} \frac{G_n(x, y)}{n} = 0, \end{aligned}$$

a contradiction.

We are now able to sharpen Theorem 3 of Chapter 1.

Theorem 3 *Let C be a finite irreducible closed set of states. Then every state in C is positive recurrent.*

Proof. The proof of this theorem is now almost immediate. Since C is a finite closed set, there is at least one positive recurrent state in C. Since C is irreducible, every state in C is positive recurrent by Theorem 2. ∎

Corollary 2 *An irreducible Markov chain having a finite number of states is positive recurrent.*

Corollary 3 *A Markov chain having a finite number of states has no null recurrent states.*

Proof. Corollary 2 follows immediately from Theorem 3. To verify Corollary 3, observe that if y is a recurrent state, then, by Theorem 4 of Chapter 1, y is contained in an irreducible closed set C of recurrent states. Since C is necessarily finite, it follows from Theorem 3 that all states in C, including y itself, are positive recurrent. Thus every recurrent state is positive recurrent, and hence there are no null recurrent states. ∎

Example 4. Consider the Markov chain described in Example 10 of Chapter 1. We have seen that 1 and 2 are transient states and that 0, 3, 4, and 5 are recurrent states. We now see that these recurrent states are necessarily positive recurrent.

2.5. Existence and uniqueness of stationary distributions

In this section we will determine which Markov chains have stationary distributions and when there is a unique such distribution. In our discussion we will need to interchange summations and limits on several occasions. This is justified by the following standard elementary result in analysis, which we state without proof.

Bounded Convergence Theorem. *Let $a(x)$, $x \in \mathscr{S}$, be nonnegative numbers having finite sum, and let $b_n(x)$, $x \in \mathscr{S}$ and $n \geq 1$, be such that $|b_n(x)| \leq 1$, $x \in \mathscr{S}$ and $n \geq 1$, and*

$$\lim_{n \to \infty} b_n(x) = b(x), \qquad x \in \mathscr{S}.$$

Then

$$\lim_{n \to \infty} \sum_x a(x) b_n(x) = \sum_x a(x) b(x).$$

Let π be a stationary distribution and let m be a positive integer. Then by (3)

$$\sum_z \pi(z) P^m(z, x) = \pi(x).$$

Summing this equation on $m = 1, 2, \ldots, n$ and dividing by n, we conclude that

$$\sum_z \pi(z) \frac{G_n(z, x)}{n} = \pi(x), \qquad x \in \mathscr{S}. \tag{34}$$

Theorem 4 *Let π be a stationary distribution. If x is a transient state or a null recurrent state, then $\pi(x) = 0$.*

Proof. If x is a transient state or a null recurrent state,

$$\lim_{n \to \infty} \frac{G_n(z, x)}{n} = 0, \qquad x \in \mathscr{S}, \tag{35}$$

as shown in Sections 2.3 and 2.4. It follows from (34), (35), and the bounded convergence theorem that

$$\pi(x) = \lim_{n \to \infty} \sum_z \pi(z) \frac{G_n(z, x)}{n} = 0,$$

as desired. ∎

It follows from this theorem that a Markov chain with no positive recurrent states does not have a stationary distribution.

Theorem 5 *An irreducible positive recurrent Markov chain has a unique stationary distribution π, given by*

$$\pi(x) = \frac{1}{m_x}, \qquad x \in \mathcal{S}. \tag{36}$$

Proof. It follows from Theorem 1 and the assumptions of this theorem that

$$\lim_{n\to\infty} \frac{G_n(z, x)}{n} = \frac{1}{m_x}, \qquad x, z \in \mathcal{S}. \tag{37}$$

Suppose π is a stationary distribution. We see from (34), (37), and the bounded convergence theorem that

$$\begin{aligned} \pi(x) &= \lim_{n\to\infty} \sum_z \pi(z) \frac{G_n(z, x)}{n} \\ &= \frac{1}{m_x} \sum_z \pi(z) = \frac{1}{m_x}. \end{aligned}$$

Thus if there is a stationary distribution, it must be given by (36).

To complete the proof of the theorem we need to show that the function $\pi(x)$, $x \in \mathcal{S}$, defined by (36) is indeed a stationary distribution. It is clearly nonnegative, so we need only show that

$$\sum_x \frac{1}{m_x} = 1 \tag{38}$$

and

$$\sum_x \frac{1}{m_x} P(x, y) = \frac{1}{m_y}, \qquad y \in \mathcal{S}. \tag{39}$$

Toward this end we observe first that

$$\sum_x P^m(z, x) = 1.$$

Summing on $m = 1, \ldots, n$ and dividing by n, we conclude that

$$\sum_x \frac{G_n(z, x)}{n} = 1, \qquad z \in \mathcal{S}. \tag{40}$$

Next we observe that by (24) of Chapter 1

$$\sum_x P^m(z, x)P(x, y) = P^{m+1}(z, y).$$

By again summing on $m = 1, \ldots, n$ and dividing by n, we conclude that

$$\sum_x \frac{G_n(z, x)}{n} P(x, y) = \frac{G_{n+1}(z, y)}{n} - \frac{P(z, y)}{n}. \tag{41}$$

If $\mathscr{S}$ is finite, we conclude from (37) and (40) that

$$1 = \lim_{n \to \infty} \sum_x \frac{G_n(z, x)}{n} = \sum_x \frac{1}{m_x},$$

i.e., that (38) holds. Similarly, we conclude that (39) holds by letting $n \to \infty$ in (41). This completes the proof of the theorem if $\mathscr{S}$ is finite.

The argument to complete the proof for $\mathscr{S}$ infinite is more complicated, since we cannot directly interchange limits and sums as we did for $\mathscr{S}$ finite (the bounded convergence theorem is not applicable). Let $\mathscr{S}_1$ be a finite subset of $\mathscr{S}$. We see from (40) that

$$\sum_{x \in \mathscr{S}_1} \frac{G_n(z, x)}{n} \le 1, \qquad z \in \mathscr{S}.$$

Since $\mathscr{S}_1$ is finite, we can let $n \to \infty$ in this inequality and conclude from (37) that

$$\sum_{x \in \mathscr{S}_1} \frac{1}{m_x} \le 1.$$

The last inequality holds for any finite subset $\mathscr{S}_1$ of $\mathscr{S}$, and hence

$$\sum_x \frac{1}{m_x} \le 1. \tag{42}$$

For if the sum of $1/m_x$ over $x \in \mathscr{S}$ exceeded 1, the sum over some finite subset of $\mathscr{S}$ would also exceed 1.

Similarly, we conclude from (41) that if $\mathscr{S}_1$ is a finite subset of $\mathscr{S}$, then

$$\sum_{x \in \mathscr{S}_1} \frac{G_n(z, x)}{n} P(x, y) \le \frac{G_{n+1}(z, y)}{n} - \frac{P(z, y)}{n}.$$

By letting $n \to \infty$ in this inequality and using (37), we obtain

$$\sum_{x \in \mathscr{S}_1} \frac{1}{m_x} P(x, y) \le \frac{1}{m_y}.$$

We conclude, as in the proof of (42), that

$$\sum_x \frac{1}{m_x} P(x, y) \le \frac{1}{m_y}, \qquad y \in \mathscr{S}. \tag{43}$$

Next we will show that equality holds in (43). It follows from (42) that the sum on y of the right side of (43) is finite. If strict inequality held for some y, it would follow by summing (43) on y that

$$\begin{aligned}\sum_y \frac{1}{m_y} &> \sum_y \left(\sum_x \frac{1}{m_x} P(x, y)\right)\\ &= \sum_x \frac{1}{m_x}\left(\sum_y P(x, y)\right)\\ &= \sum_x \frac{1}{m_x},\end{aligned}$$

which is a contradiction. This proves that equality holds in (43), i.e., that (39) holds.

Set

$$c = \frac{1}{\sum_x \frac{1}{m_x}}.$$

Then by (39)

$$\pi(x) = \frac{c}{m_x}, \qquad x \in \mathscr{S},$$

defines a stationary distribution. Thus by the first part of the proof of this theorem

$$\frac{c}{m_x} = \frac{1}{m_x},$$

and hence $c = 1$. This proves that (38) holds and completes the proof of the theorem. ∎

From Theorems 4 and 5 we immediately obtain

Corollary 4 *An irreducible Markov chain is positive recurrent if and only if it has a stationary distribution.*

Example 5. Consider an irreducible birth and death chain on the nonnegative integers. Find necessary and sufficient conditions for the chain to be

(a) positive recurrent,
(b) null recurrent,
(c) transient.

From Section 2.2.1 we see that the chain has a stationary distribution if and only if

$$\sum_{x=1}^{\infty} \frac{p_0 \cdots p_{x-1}}{q_1 \cdots q_x} < \infty. \tag{44}$$

Thus (44) is necessary and sufficient for the chain to be positive recurrent. We saw in Section 1.7 that

$$\sum_{x=1}^{\infty} \frac{q_1 \cdots q_x}{p_1 \cdots p_x} < \infty \tag{45}$$

is a necessary and sufficient condition for the chain to be transient. For the chain to be null recurrent, it is necessary and sufficient that (44) and (45) both fail to hold. Thus

$$\sum_{x=1}^{\infty} \frac{q_1 \cdots q_x}{p_1 \cdots p_x} = \infty \qquad \text{and} \qquad \sum_{x=1}^{\infty} \frac{p_0 \cdots p_{x-1}}{q_1 \cdots q_x} = \infty \tag{46}$$

are necessary and sufficient conditions for the chain to be null recurrent.

As an immediate consequence of Corollary 2 and Theorem 5 we obtain

Corollary 5 *If a Markov chain having a finite number of states is irreducible, it has a unique stationary distribution.*

Recall that $N_n(x)$ denotes the number of visits to x during times $m = 1, \ldots, n$. By combining Corollary 1 and Theorem 5 we get

Corollary 6 *Let X_n, $n \geq 0$, be an irreducible positive recurrent Markov chain having stationary distribution π. Then with probability one*

$$\lim_{n \to \infty} \frac{N_n(x)}{n} = \pi(x), \qquad x \in \mathscr{S}. \tag{47}$$

2.5.1. Reducible chains. Let π be a distribution on $\mathscr{S}$, i.e., let $\pi(x)$, $x \in \mathscr{S}$, be nonnegative numbers adding to one, and let C be a subset of $\mathscr{S}$. We say that π is *concentrated* on C if

$$\pi(x) = 0, \qquad x \notin C.$$

By essentially the same argument used to prove Theorem 5 we can obtain a somewhat more general result.

Theorem 6 *Let C be an irreducible closed set of positive recurrent states. Then the Markov chain has a unique stationary distribution π concentrated on C. It is given by*

$$\pi(x) = \begin{cases} \dfrac{1}{m_x}, & x \in C, \\ 0, & \text{elsewhere.} \end{cases} \tag{48}$$

Suppose C_0 and C_1 are two distinct irreducible closed sets of positive recurrent states of a Markov chain. It follows from Theorem 6 that the Markov chain has a stationary distribution π_0 concentrated on C_0 and a different stationary distribution π_1 concentrated on C_1. Moreover, the distributions π_α defined for $0 \le \alpha \le 1$ by

$$\pi_\alpha(x) = (1 - \alpha)\pi_0(x) + \alpha\pi_1(x), \qquad x \in \mathscr{S},$$

are distinct stationary distributions (see Exercise 5).

By combining Theorems 4–6 and their consequences, we obtain

***Corollary** 7 Let $\mathscr{S}_P$ denote the positive recurrent states of a Markov chain.*

(i) *If $\mathscr{S}_P$ is empty, the chain has no stationary distributions.*
(ii) *If $\mathscr{S}_P$ is a nonempty irreducible set, the chain has a unique stationary distribution.*
(iii) *If $\mathscr{S}_P$ is nonempty but not irreducible, the chain has an infinite number of distinct stationary distributions.*

Consider now a Markov chain having a finite number of states. Then every recurrent state is positive recurrent and there is at least one such state. There are two possibilities: either the set $\mathscr{S}_R$ of recurrent states is irreducible and there is a unique stationary distribution, or $\mathscr{S}_R$ can be decomposed into two or more irreducible closed sets and there is an infinite number of distinct stationary distributions. The latter possibility holds for a Markov chain on $\mathscr{S} = \{0, 1, \ldots, d\}$ in which $d > 0$ and 0 and d are both absorbing states. The gambler's ruin chain on $\{0, 1, \ldots, d\}$ and the genetics model in Example 7 of Chapter 1 are of this type. For such a chain any distribution π_α, $0 \le \alpha \le 1$, of the form

$$\pi_\alpha(x) = \begin{cases} 1 - \alpha, & x = 0, \\ \alpha, & x = d, \\ 0, & \text{elsewhere}, \end{cases}$$

is a stationary distribution.

Example 6. Consider the Markov chain introduced in Example 10 of Chapter 1. Find the stationary distribution concentrated on each of the irreducible closed sets.

We saw in Section 1.6 that the set of recurrent states for this chain is decomposed into the absorbing state 0 and the irreducible closed set $\{3, 4, 5\}$. Clearly the unique stationary distribution concentrated on $\{0\}$ is given by $\boldsymbol{\pi}_0 = (1, 0, 0, 0, 0, 0)$. To find the unique stationary distribution concentrated on {3, 4, 5}, we must find nonnegative numbers $\pi(3)$, $\pi(4)$, and $\pi(5)$ summing to one and satisfying the three equations

$$\frac{\pi(3)}{6} + \frac{\pi(4)}{2} + \frac{\pi(5)}{4} = \pi(3)$$

$$\frac{\pi(3)}{3} = \pi(4)$$

$$\frac{\pi(3)}{2} + \frac{\pi(4)}{2} + \frac{3\pi(5)}{4} = \pi(5).$$

From the first two of these equations we find that $\pi(4) = \pi(3)/3$ and $\pi(5) = 8\pi(3)/3$. Thus

$$\pi(3)(1 + \tfrac{1}{3} + \tfrac{8}{3}) = 1,$$

from which we conclude that

$$\pi(3) = \tfrac{1}{4}, \qquad \pi(4) = \tfrac{1}{12}, \qquad \text{and} \qquad \pi(5) = \tfrac{2}{3}.$$

Consequently

$$\pi_1 = (0, 0, 0, \tfrac{1}{4}, \tfrac{1}{12}, \tfrac{2}{3})$$

is the stationary distribution concentrated on {3, 4, 5}.

2.6. Queuing chain

Consider the queuing chain introduced in Example 5 of Chapter 1. Recall that the number of customers arriving in unit time has density f and mean μ. Suppose that the chain is irreducible, which means that $f(0) > 0$ and $f(0) + f(1) < 1$ (see Exercise 37 of Chapter 1). In Chapter 1 we saw that the chain is recurrent if $\mu \leq 1$ and transient if $\mu > 1$. In Section 2.6.1 we will show that in the recurrent case

$$m_0 = \frac{1}{1 - \mu}. \tag{49}$$

It follows from (49) that if $\mu < 1$, then $m_0 < \infty$ and hence 0 is a positive recurrent state. Thus by irreducibility the chain is positive recurrent. On the other hand, if $\mu = 1$, then $m_0 = \infty$ and hence 0 is a null recurrent state. We conclude that the queuing chain is null recurrent in this case. Therefore *an irreducible queuing chain is positive recurrent if* $\mu < 1$ *and null recurrent if* $\mu = 1$, *and transient if* $\mu > 1$.

***2.6.1.** Proof. We will now verify (49). We suppose throughout the proof of this result that $f(0) > 0$, $f(0) + f(1) < 1$ and $\mu \leq 1$, so that the chain is irreducible and recurrent. Consider such a chain starting at the positive integer x. Then T_{x-1} denotes the time to go from state x to state $x - 1$, and $T_{y-1} - T_y$, $1 \leq y \leq x - 1$, denotes the time to go from state y to state $y - 1$. Since the queuing chain goes at most one step to the left at a time, the Markov property insures that the random variables

$$T_{x-1},\ T_{x-2} - T_{x-1}, \ldots, T_0 - T_1$$

are independent. These random variables are identically distributed; for each of them is distributed as

$$\min\,(n > 0\colon \xi_1 + \cdots + \xi_n = n - 1),$$

i.e., as the smallest positive integer n such that the number of customers served by time n is one more than the number of new customers arriving by time n.

Let $G(t)$, $0 \leq t \leq 1$, denote the probability generation function of the time to go from state 1 to state 0. Then

$$G(t) = \sum_{n=1}^{\infty} t^n P_1(T_0 = n). \tag{50}$$

The probability generating function of the sum of independent nonnegative integer-valued random variables is the product of their respective probability generating functions. If the chain starts at x, then

$$T_0 = T_{x-1} + (T_{x-2} - T_{x-1}) + \cdots + (T_0 - T_1)$$

is the sum of x independent random variables each having probability generating function $G(t)$. Thus the probability generating function of T_0 is $(G(t))^x$; that is,

$$(G(t))^x = \sum_{n=1}^{\infty} t^n P_x(T_0 = n). \tag{51}$$

We will now show that

$$G(t) = t\Phi(G(t)), \qquad 0 \leq t \leq 1, \tag{52}$$

where Φ denotes the probability generating function of f. To verify (52) we rewrite (50) as

$$G(t) = \sum_{n=0}^{\infty} t^{n+1} P_1(T_0 = n + 1) = tP(1, 0) + t \sum_{n=1}^{\infty} t^n P_1(T_0 = n + 1).$$

* This material is optional and can be omitted with no loss of continuity.

By using successively (29) of Chapter 1, (51) of this chapter, and the formula $P(1, y) = f(y)$, $y \geq 0$, we find that

$$\begin{aligned} G(t) &= tP(1, 0) + t \sum_{n=1}^{\infty} t^n \sum_{y \neq 0} P(1, y)P_y(T_0 = n) \\ &= tP(1, 0) + t \sum_{y \neq 0} P(1, y) \sum_{n=1}^{\infty} t^n P_y(T_0 = n) \\ &= tP(1, 0) + t \sum_{y \neq 0} P(1, y)(G(t))^y \\ &= t\left[f(0) + \sum_{y \neq 0} f(y)(G(t))^y\right] \\ &= t\Phi(G(t)). \end{aligned}$$

For $0 \leq t < 1$ we can differentiate both sides of (52) and obtain

$$G'(t) = \Phi(G(t)) + tG'(t)\Phi'(G(t)).$$

Solving for $G'(t)$ we find that

$$G'(t) = \frac{\Phi(G(t))}{1 - t\Phi'(G(t))}, \qquad 0 \leq t < 1. \tag{53}$$

Now $G(t) \to 1$ and $\Phi(t) \to 1$ as $t \to 1$ and

$$\begin{aligned} \lim_{t \to 1} \Phi'(t) &= \lim_{t \to 1} \sum_{x=1}^{\infty} x f(x) t^{x-1} \\ &= \sum_{x=1}^{\infty} x f(x) = \mu. \end{aligned}$$

By letting $t \to 1$ in (53) we see that

$$\lim_{t \to 1} G'(t) = \frac{1}{1 - \mu}. \tag{54}$$

By definition

$$G(t) = \sum_{n=1}^{\infty} P_1(T_0 = n)t^n.$$

But since $P(1, x) = P(0, x)$, $x \geq 0$, it follows from (29) of Chapter 1 that the distribution of T_0 for a queuing chain starting in state 1 is the same as that for a chain starting in state 0. Consequently,

$$G(t) = \sum_{n=1}^{\infty} P_0(T_0 = n)t^n,$$

and hence

$$\lim_{t \to 1} G'(t) = \lim_{t \to 1} \sum_{n=1}^{\infty} nP_0(T_0 = n)t^{n-1}$$
$$= \sum_{n=1}^{\infty} nP_0(T_0 = n)$$
$$= E_0(T_0) = m_0.$$

It now follows from (54) that (49) holds. ▮

2.7. Convergence to the stationary distribution

We have seen earlier in this chapter that if X_n, $n \geq 0$, is an irreducible positive recurrent Markov chain having π as its stationary distribution, then

$$\lim_{n \to \infty} \frac{1}{n} \sum_{m=1}^{n} P^m(x, y) = \lim_{n \to \infty} \frac{G_n(x, y)}{n} = \pi(y), \qquad x, y \in \mathscr{S}.$$

In this section we will see when the stronger result

$$\lim_{n \to \infty} P^n(x, y) = \pi(y), \qquad x, y \in \mathscr{S},$$

holds and what happens when it fails to hold.

The positive integer d is said to be a *divisor* of the positive integer n if n/d is an integer. If I is a nonempty set of positive integers, the *greatest common divisor of I*, denoted by g.c.d. I, is defined to be the largest integer d such that d is a divisor of every integer in I. It follows immediately that

$$1 \leq \text{g.c.d. } I \leq \min\,(n : n \in I).$$

In particular, if $1 \in I$, then g.c.d. $I = 1$. The greatest common divisor of the set of even positive integers is 2.

Let x be a state of a Markov chain such that $P^n(x, x) > 0$ for some $n \geq 1$, i.e., such that $\rho_{xx} = P_x(T_x < \infty) > 0$. We define its *period* d_x by

$$d_x = \text{g.c.d. } \{n \geq 1 : P^n(x, x) > 0\}.$$

Then

$$1 \leq d_x \leq \min\,(n \geq 1 : P^n(x, x) > 0).$$

If $P(x, x) > 0$, then $d_x = 1$.

If x and y are two states, each of which leads to the other, then $d_x = d_y$. For let n_1 and n_2 be positive integers such that

$$P^{n_1}(x, y) > 0 \qquad \text{and} \qquad P^{n_2}(y, x) > 0.$$

Then

$$P^{n_1+n_2}(x, x) \geq P^{n_1}(x, y)P^{n_2}(y, x) > 0,$$

and hence d_x is a divisor of $n_1 + n_2$. If $P^n(y, y) > 0$, then

$$P^{n_1+n+n_2}(x, x) \geq P^{n_1}(x, y)P^n(y, y)P^{n_2}(y, x) > 0,$$

so that d_x is a divisor of $n_1 + n + n_2$. Since d_x is a divisor of $n_1 + n_2$, it must be a divisor of n. Thus d_x is a divisor of all numbers in the set $\{n \geq 1: P^n(y, y) > 0\}$. Since d_y is the largest such divisor, we conclude that $d_x \leq d_y$. Similarly $d_y \leq d_x$, and hence $d_x = d_y$.

We have shown, in other words, that the states in an irreducible Markov chain have common period d. We say that the chain is *periodic with period* d if $d > 1$ and *aperiodic* if $d = 1$. A simple sufficient condition for an irreducible Markov chain to be aperiodic is that $P(x, x) > 0$ for some $x \in \mathscr{S}$. Since $P(0, 0) = f(0) > 0$ for an irreducible queuing chain, such a chain is necessarily aperiodic.

Example 7. Determine the period of an irreducible birth and death chain.

If some $r_x > 0$, then $P(x, x) = r_x > 0$, and the birth and death chain is aperiodic. In particular, the modified Ehrenfest chain in Example 3 is aperiodic.

Suppose $r_x = 0$ for all x. Then in one transition the state of the chain changes either from an odd numbered state to an even numbered state or from an even numbered state to an odd numbered state. In particular, a chain can return to its initial state only after an even number of transitions. Thus the period of the chain is 2 or a multiple of 2. Since

$$P^2(0, 0) = p_0 q_1 > 0,$$

we conclude that the chain is periodic with period 2. In particular, the Ehrenfest chain introduced in Example 2 of Chapter 1 is periodic with period 2.

Theorem 7 *Let X_n, $n \geq 0$, be an irreducible positive recurrent Markov chain having stationary distribution π. If the chain is aperiodic,*

$$\lim_{n\to\infty} P^n(x, y) = \pi(y), \qquad x, y \in \mathscr{S}. \tag{55}$$

If the chain is periodic with period d, then for each pair x, y of states in $\mathscr{S}$ there is an integer r, $0 \leq r < d$, such that $P^n(x, y) = 0$ unless $n = md + r$ for some nonnegative integer m, and

$$\lim_{m\to\infty} P^{md+r}(x, y) = d\pi(y). \tag{56}$$

For an illustration of the second half of this theorem, consider an irreducible positive recurrent birth and death chain which is periodic with period 2. If $y - x$ is even, then $P^{2m+1}(x, y) = 0$ for all $m \geq 0$ and

$$\lim_{m \to \infty} P^{2m}(x, y) = 2\pi(y).$$

If $y - x$ is odd, then $P^{2m}(x, y) = 0$ for all $m \geq 1$ and

$$\lim_{m \to \infty} P^{2m+1}(x, y) = 2\pi(y).$$

We will prove this theorem in an appendix to this chapter, which can be omitted with no loss of continuity.

Example 8. Determine the asymptotic behavior of the matrix P^n for the transition matrix P

(a) from Example 3,

(b) from Example 2.

(a) The transition matrix P from Example 3 corresponds to an aperiodic irreducible Markov chain on $\{0, 1, 2, 3\}$ having the stationary distribution given by

$$\pi(0) = \tfrac{1}{8}, \qquad \pi(1) = \tfrac{3}{8}, \qquad \pi(2) = \tfrac{3}{8}, \qquad \text{and} \qquad \pi(3) = \tfrac{1}{8}.$$

It follows from Theorem 7 that for n large

$$P^n \doteq \begin{bmatrix} \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8} \\ \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8} \\ \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8} \\ \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8} \end{bmatrix}.$$

(b) The transition matrix P from Example 2 corresponds to a periodic irreducible Markov chain on $\{0, 1, 2, 3\}$ having period 2 and the same stationary distribution as the chain in Example 3. From the discussion following the statement of Theorem 7, we conclude that for n large and even

$$P^n \doteq \begin{bmatrix} \frac{1}{4} & 0 & \frac{3}{4} & 0 \\ 0 & \frac{3}{4} & 0 & \frac{1}{4} \\ \frac{1}{4} & 0 & \frac{3}{4} & 0 \\ 0 & \frac{3}{4} & 0 & \frac{1}{4} \end{bmatrix},$$

while for n large and odd

$$P^n \doteq \begin{bmatrix} 0 & \frac{3}{4} & 0 & \frac{1}{4} \\ \frac{1}{4} & 0 & \frac{3}{4} & 0 \\ 0 & \frac{3}{4} & 0 & \frac{1}{4} \\ \frac{1}{4} & 0 & \frac{3}{4} & 0 \end{bmatrix}.$$

APPENDIX

2.8. Proof of convergence

We will first prove Theorem 7 in the aperiodic case. Consider an aperiodic, irreducible, positive recurrent Markov chain having transition function P, state space $\mathscr{S}$, and stationary distribution π. We will now verify that the conclusion of Theorem 7 holds for such a chain.

Choose $a \in \mathscr{S}$ and let I be the set of positive integers defined by

$$I = \{n > 0 : P^n(a, a) > 0\}.$$

Then

(i) g.c.d. $I = 1$;
(ii) if $m \in I$ and $n \in I$, then $m + n \in I$.

Property (ii) follows from the inequality

$$P^{m+n}(a, a) \geq P^m(a, a)P^n(a, a).$$

Properties (i) and (ii) imply that there is a positive integer n_1 such that $n \in I$ for all $n \geq n_1$. For completeness we will prove this number theoretic result in Section 2.8.2. Using this result we conclude that $P^n(a, a) > 0$ for $n \geq n_1$.

Let x and y be any pair of states in $\mathscr{S}$. Since the chain is irreducible, there exist positive integers n_2 and n_3 such that

$$P^{n_2}(x, a) > 0 \qquad \text{and} \qquad P^{n_3}(a, y) > 0.$$

Then for $n \geq n_1$

$$P^{n_2+n+n_3}(x, y) \geq P^{n_2}(x, a)P^n(a, a)P^{n_3}(a, y) > 0.$$

We have shown, in other words, that for every pair x, y of states in $\mathscr{S}$ there is a positive integer n_0 such that

$$P^n(x, y) > 0, \qquad n \geq n_0. \tag{57}$$

Set

$$\mathscr{S}^2 = \{(x, y) : x \in \mathscr{S} \text{ and } y \in \mathscr{S}\}.$$

Then $\mathscr{S}^2$ is the set of ordered pairs of elements in $\mathscr{S}$. We will consider a Markov chain (X_n, Y_n) having state space $\mathscr{S}^2$ and transition function P_2 defined by

$$P_2((x_0, y_0), (x, y)) = P(x_0, x)P(y_0, y).$$

It follows that X_n, $n \geq 0$, and Y_n, $n \geq 0$, are each Markov chains having transition function P, and the successive transitions of the X_n chain and the Y_n chain are chosen independently of each other.

We will now develop properties of the Markov chain (X_n, Y_n). In particular, we will show that this chain is an aperiodic, irreducible, positive recurrent Markov chain. We will then use this chain to verify the conclusion of the theorem.

Choose $(x_0, y_0) \in \mathscr{S}^2$ and $(x, y) \in \mathscr{S}^2$. By (57) there is an $n_0 > 0$ such that

$$P^n(x_0, x) > 0 \qquad \text{and} \qquad P^n(y_0, y) > 0, \qquad n \geq n_0.$$

Then

$$(58) \qquad P_2^n((x_0, y_0), (x, y)) = P^n(x_0, x)P^n(y_0, y) > 0, \qquad n \geq n_0.$$

We conclude from (58) that the chain is both irreducible and aperiodic.

The distribution π_2 on $\mathscr{S}^2$ defined by $\pi_2(x_0, y_0) = \pi(x_0)\pi(y_0)$ is a stationary distribution. For

$$\begin{aligned}\sum_{(x_0,y_0) \in \mathscr{S}^2} & \pi_2(x_0, y_0)P_2((x_0, y_0), (x, y)) \\ &= \sum_{x_0 \in \mathscr{S}} \sum_{y_0 \in \mathscr{S}} \pi(x_0)\pi(y_0)P(x_0, x)P(y_0, y) \\ &= \left(\sum_{x_0 \in \mathscr{S}} \pi(x_0)P(x_0, x)\right)\left(\sum_{y_0 \in \mathscr{S}} \pi(y_0)P(y_0, y)\right) \\ &= \pi(x)\pi(y) = \pi_2(x, y).\end{aligned}$$

Thus the chain on $\mathscr{S}^2$ is positive recurrent; in particular, it is recurrent.

Set

$$T = \min (n > 0 : X_n = Y_n).$$

Choose $a \in \mathscr{S}$. Since the (X_n, Y_n) chain is recurrent,

$$T_{(a,a)} = \min (n > 0 : (X_n, Y_n) = (a, a))$$

is finite with probability one. Clearly $T \leq T_{(a,a)}$, and hence T is finite with probability one.

For any $n \geq 1$ (regardless of the distribution of (X_0, Y_0))

$$(59) \qquad P(X_n = y, T \leq n) = P(Y_n = y, T \leq n), \qquad y \in \mathscr{S}.$$

This formula is intuitively reasonable since the two chains are indistinguishable for $n \geq T$. To make this argument precise, we choose $1 \leq m \leq n$. Then for $z \in \mathscr{S}$

$$\begin{aligned}(60) \qquad P(X_n = y \mid T = m, X_m = Y_m = z)& \\ &= P(Y_n = y \mid T = m, X_m = Y_m = z),\end{aligned}$$

since both conditional probabilities equal $P^{n-m}(z, y)$. Now the event $\{T \leq n\}$ is the union of the disjoint events

$$\{T = m, X_m = Y_m = z\}, \qquad 1 \leq m \leq n \quad \text{and} \quad z \in \mathscr{S},$$

so it follows from (60) and Exercise 4(d) of Chapter 1 that

$$P(X_n = y \mid T \le n) = P(Y_n = y \mid T \le n)$$

and hence that (59) holds.

Equation (59) implies that

$$\begin{aligned} P(X_n = y) &= P(X_n = y, T \le n) + P(X_n = y, T > n) \\ &= P(Y_n = y, T \le n) + P(X_n = y, T > n) \\ &\le P(Y_n = y) + P(T > n) \end{aligned}$$

and similarly that

$$P(Y_n = y) \le P(X_n = y) + P(T > n).$$

Therefore for $n \ge 1$

(61) $$|P(X_n = y) - P(Y_n = y)| \le P(T > n), \qquad y \in \mathscr{S}.$$

Since T is finite with probability one,

(62) $$\lim_{n\to\infty} P(T > n) = 0.$$

We conclude from (61) and (62) that

(63) $$\lim_{n\to\infty} (P(X_n = y) - P(Y_n = y)) = 0, \qquad y \in \mathscr{S}.$$

Using (63), we can easily complete the proof of Theorem 7. Choose $x \in \mathscr{S}$ and let the initial distribution of (X_n, Y_n) be such that $P(X_0 = x) = 1$ and

$$P(Y_0 = y_0) = \pi(y_0), \qquad y_0 \in \mathscr{S}.$$

Since X_n, $n \ge 0$, and Y_n, $n \ge 0$, are each Markov chains with transition function P, we see that

(64) $$P(X_n = y) = P^n(x, y), \qquad y \in \mathscr{S},$$

and

(65) $$P(Y_n = y) = \pi(y), \qquad y \in \mathscr{S}.$$

Thus by (63)–(65)

$$\lim_{n\to\infty} (P^n(x, y) - \pi(y)) = \lim_{n\to\infty} (P(X_n = y) - P(Y_n = y)) = 0,$$

and hence the conclusion of Theorem 7 holds.

2.8.1. Periodic case. We first consider a slight extension of Theorem 7 in the aperiodic case. Let C be an irreducible closed set of positive recurrent states such that each state in C has period 1, and let π

be the unique stationary distribution concentrated on C. By looking at the Markov chain restricted to C, we conclude that

$$\lim_{n\to\infty} P^n(x, y) = \pi(y) = \frac{1}{m_y}, \qquad x, y \in C.$$

In particular, if y is any positive recurrent state having period 1, then by letting C be the irreducible closed set containing y, we see that

$$\lim_{n\to\infty} P^n(y, y) = \frac{1}{m_y}. \tag{66}$$

We now proceed with the proof of Theorem 7 in the periodic case. Let X_n, $n \geq 0$, be an irreducible positive recurrent Markov chain which is periodic with period $d > 1$. Set $Y_m = X_{md}$, $m \geq 0$. Then Y_m, $m \geq 0$, is a Markov chain having transition function $Q = P^d$. Choose $y \in \mathscr{S}$. Then

$$\begin{aligned} \text{g.c.d. } \{m \mid Q^m(y, y) > 0\} &= \text{g.c.d. } \{m \mid P^{md}(y, y) > 0\} \\ &= \frac{1}{d} \text{ g.c.d. } \{n \mid P^n(y, y) > 0\} \\ &= 1. \end{aligned}$$

Thus all states have period 1 with respect to the Y_m chain.

Let the X_n chain and hence also the Y_m chain start at y. Since the X_n chain first returns to y at some multiple of d, it follows that the expected return time to y for the Y_m chain is $d^{-1}m_y$, where m_y is the expected return time to y for the X_n chain. In particular, y is a positive recurrent state for a Markov chain having transition function Q. By applying (66) to this transition function we conclude that

$$\lim_{m\to\infty} Q^m(y, y) = \frac{d}{m_y} = d\pi(y),$$

and thus that

$$\lim_{m\to\infty} P^{md}(y, y) = d\pi(y), \qquad y \in \mathscr{S}. \tag{67}$$

Let x and y be any pair of states in $\mathscr{S}$ and set

$$r_1 = \min (n : P^n(x, y) > 0).$$

Then, in particular, $P^{r_1}(x, y) > 0$. We will show that $P^n(x, y) > 0$ only if $n - r_1$ is an integral multiple of d.

Choose n_1 such that $P^{n_1}(y, x) > 0$. Then

$$P^{r_1+n_1}(y, y) \geq P^{n_1}(y, x)P^{r_1}(x, y) > 0,$$

and hence $r_1 + n_1$ is an integral multiple of d. If $P^n(x, y) > 0$, then by the same argument $n + n_1$ is an integral multiple of d, and therefore so is $n - r_1$. Thus, $n = kd + r_1$ for some nonnegative integer k.

There is a nonnegative integer m_1 such that $r_1 = m_1 d + r$, where $0 \le r < d$. We conclude that

$$P^n(x, y) = 0 \qquad \text{unless} \qquad n = md + r \tag{68}$$

for some nonnegative integer m. It follows from (68) and from (28) of Chapter 1 that

$$P^{md+r}(x, y) = \sum_{k=0}^{m} P_x(T_y = kd + r)P^{(m-k)d}(y, y). \tag{69}$$

Set

$$a_m(k) = \begin{cases} P^{(m-k)d}(y, y), & 0 \le k \le m, \\ 0, & k > m. \end{cases}$$

Then by (67) for each fixed k

$$\lim_{m \to \infty} a_m(k) = d\pi(y).$$

We can apply the bounded convergence theorem (with $\mathscr{S}$ replaced by $\{0, 1, 2, \ldots\}$) to conclude from (69) that

$$\begin{aligned} \lim_{m \to \infty} P^{md+r}(x, y) &= d\pi(y) \sum_{k=0}^{\infty} P_x(T_y = kd + r) \\ &= d\pi(y) P_x(T_y < \infty) \\ &= d\pi(y), \end{aligned}$$

and hence that (56) holds. This completes the proof of Theorem 7. ∎

2.8.2. A result from number theory. Let I be a nonempty set of positive integers such that

(i) g.c.d. $I = 1$;

(ii) if m and n are in I, then $m + n$ is in I.

Then there is an n_0 such that $n \in I$ for all $n \ge n_0$.

We will first prove that I contains two consecutive integers. Suppose otherwise. Then there is an integer $k \ge 2$ and an $n_1 \in I$ such that $n_1 + k \in I$ and any two distinct integers in I differ by at least k. It follows from property (i) that there is an $n \in I$ such that k is not a divisor of n. We can write

$$n = mk + r,$$

where m is a nonnegative integer and $0 < r < k$. It follows from property (ii) that $(m + 1)(n_1 + k)$ and $n + (m + 1)n_1$ are each in I. Their difference is

$$(m + 1)(n_1 + k) - n - (m + 1)n_1 = k + mk - n = k - r,$$

which is positive and smaller than k. This contradicts the definition of k.

We have shown that I contains two consecutive integers, say n_1 and $n_1 + 1$. Let $n \geq n_1^2$. Then there are nonnegative integers m and r such that $0 \leq r < n_1$ and

$$n - n_1^2 = mn_1 + r.$$

Thus

$$n = r(n_1 + 1) + (n_1 - r + m)n_1,$$

which is in I by property (ii). This shows that $n \in I$ for all

$$n \geq n_0 = n_1^2. \qquad \blacksquare$$

Exercises

1 Consider a Markov chain having state space $\{0, 1, 2\}$ and transition matrix

$$\begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{bmatrix} .4 & .4 & .2 \\ .3 & .4 & .3 \\ .2 & .4 & .4 \end{bmatrix} \end{array}.$$

Show that this chain has a unique stationary distribution π and find π.

2 Consider a Markov chain having transition function P such that $P(x, y) = \alpha_y$, $x \in \mathscr{S}$ and $y \in \mathscr{S}$, where the α_y's are constants. Show that the chain has a unique stationary distribution π, given by $\pi(y) = \alpha_y$, $y \in \mathscr{S}$.

3 Let π be a stationary distribution of a Markov chain. Show that if $\pi(x) > 0$ and x leads to y, then $\pi(y) > 0$.

4 Let π be a stationary distribution of a Markov chain. Suppose that y and z are two states such that for some constant c

$$P(x, y) = cP(x, z), \qquad x \in \mathscr{S}.$$

Show that $\pi(y) = c\pi(z)$.

5 Let π_0 and π_1 be distinct stationary distributions for a Markov chain.

(a) Show that for $0 \leq \alpha \leq 1$, the function π_α defined by

$$\pi_\alpha(x) = (1 - \alpha)\pi_0(x) + \alpha\pi_1(x), \qquad x \in \mathscr{S},$$

is a stationary distribution.

(b) Show that distinct values of α determine distinct stationary distributions π_α. *Hint:* Choose $x_0 \in \mathscr{S}$ such that $\pi_0(x_0) \neq \pi_1(x_0)$ and show that $\pi_\alpha(x_0) = \pi_\beta(x_0)$ implies that $\alpha = \beta$.

6 Consider a birth and death chain on the nonnegative integers and suppose that $p_0 = 1, p_x = p > 0$ for $x \geq 1$, and $q_x = q = 1 - p > 0$ for $x \geq 1$. Find the stationary distribution when it exists.

7 (a) Find the stationary distribution of the Ehrenfest chain.
(b) Find the mean and variance of this distribution.

8 For general d, find the transition function of the modified Ehrenfest chain introduced in Example 3, and show that this chain has the same stationary distribution as does the original Ehrenfest chain.

9 Find the stationary distribution of the birth and death chain described in Exercise 2 of Chapter 1. *Hint:* Use the formula

$$\binom{d}{0}^2 + \cdots + \binom{d}{d}^2 = \binom{2d}{d}.$$

10 Let X_n, $n \geq 0$, be a positive recurrent irreducible birth and death chain, and suppose that X_0 has the stationary distribution π. Show that

$$P(X_0 = y \mid X_1 = x) = P(x, y), \qquad x, y \in \mathscr{S}.$$

Hint: Use the definition of π_x given by (9).

11 Let X_n, $n \geq 0$, be the Markov chain introduced in Section 2.2.2. Show that if X_0 has a Poisson distribution with parameter t, then X_n has a Poisson distribution with parameter

$$tp^n + \frac{\lambda}{q}(1 - p^n).$$

12 Let X_n, $n \geq 0$, be as in Exercise 11. Show that

$$E_x(X_n) = xp^n + \frac{\lambda}{q}(1 - p^n).$$

Hint: Use the result of Exercise 11 and equate coefficients of t^x in the appropriate power series.

13 Let X_n, $n \geq 0$, be as in Exercise 11 and suppose that X_0 has the stationary distribution. Use the result of Exercise 12 to find $\operatorname{cov}(X_m, X_{m+n})$, $m \geq 0$ and $n \geq 0$.

14 Consider a Markov chain on the nonnegative integers having transition function P given by $P(x, x + 1) = p$ and $P(x, 0) = 1 - p$, where $0 < p < 1$. Show that this chain has a unique stationary distribution π and find π.

15 The transition function of a Markov chain is called *doubly stochastic* if

$$\sum_{x \in \mathscr{S}} P(x, y) = 1, \qquad y \in \mathscr{S}.$$

What is the stationary distribution of an irreducible Markov chain having $d < \infty$ states and a doubly stochastic transition function?

16 Consider an irreducible Markov chain having finite state space $\mathscr{S}$, transition function P such that $P(x, x) = 0$, $x \in \mathscr{S}$ and stationary distribution π. Let p_x, $x \in \mathscr{S}$, be such that $0 < p_x < 1$, and let $Q(x, y)$, $x \in \mathscr{S}$ and $y \in \mathscr{S}$, be defined by

$$Q(x, x) = 1 - p_x$$

and

$$Q(x, y) = p_x P(x, y), \qquad y \neq x.$$

Show that Q is the transition function of an irreducible Markov chain having state space $\mathscr{S}$ and stationary distribution π', defined by

$$\pi'(x) = \frac{p_x^{-1}\pi(x)}{\sum_{y \in \mathscr{S}} p_y^{-1}\pi(y)}, \qquad x \in \mathscr{S}.$$

The interpretation of the chain with transition function Q is that starting from x, it has probability $1 - p_x$ of remaining in x and probability p_x of jumping according to the transition function P.

17 Consider the Ehrenfest chain. Suppose that initially all of the balls are in the second box. Find the expected amount of time until the system returns to that state. *Hint:* Use the result of Exercise 7(a).

18 A particle moves according to a Markov chain on $\{1, 2, \ldots, c + d\}$, where c and d are positive integers. Starting from any one of the first c states, the particle jumps in one transition to a state chosen uniformly from the last d states; starting from any of the last d states, the particle jumps in one transition to a state chosen uniformly from the first c states.

(a) Show that the chain is irreducible.
(b) Find the stationary distribution.

19 Consider a Markov chain having the transition matrix given by Exercise 19 of Chapter 1.

(a) Find the stationary distribution concentrated on each of the irreducible closed sets.
(b) Find $\lim_{n \to \infty} G_n(x, y)/n$.

20 Consider a Markov chain having transition matrix as in Exercise 20 of Chapter 1.

(a) Find the stationary distribution concentrated on each of the irreducible closed sets.
(b) Find $\lim_{n \to \infty} G_n(x, y)/n$.

21 Let X_n, $n \geq 0$, be the Ehrenfest chain with $d = 4$ and $X_0 = 0$.
(a) Find the approximate distribution of X_n for n large and even.
(b) Find the approximate distribution of X_n for n large and odd.

22 Consider a Markov chain on $\{0, 1, 2\}$ having transition matrix

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} \end{array}.$$

(a) Show that the chain is irreducible.
(b) Find the period.
(c) Find the stationary distribution.

23 Consider a Markov chain on $\{0, 1, 2, 3, 4\}$ having transition matrix

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 0 & \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & \frac{3}{4} \\ 0 & 0 & 0 & \frac{1}{4} & \frac{3}{4} \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}.$$

(a) Show that the chain is irreducible.
(b) Find the period.
(c) Find the stationary distribution.

3 Markov Pure Jump Processes

Consider again a system that at any time can be in one of a finite or countably infinite set $\mathscr{S}$ of states. We call $\mathscr{S}$ the *state space* of the system. In Chapters 1 and 2 we studied the behavior of such systems at integer times. In this chapter we will study the behavior of such systems over all times $t \geq 0$.

3.1. Construction of jump processes

Consider a system starting in state x_0 at time 0. We suppose that the system remains in state x_0 until some positive time τ_1, at which time the system jumps to a new state $x_1 \neq x_0$. We allow the possibility that the system remains permanently in state x_0, in which case we set $\tau_1 = \infty$. If τ_1 is finite, upon reaching x_1 the system remains there until some time $\tau_2 > \tau_1$ when it jumps to state $x_2 \neq x_1$. If the system never leaves x_1, we set $\tau_2 = \infty$. This procedure is repeated indefinitely. If some $\tau_m = \infty$, we set $\tau_n = \infty$ for $n > m$.

Let $X(t)$ denote the state of the system at time t, defined by

$$X(t) = \begin{cases} x_0, & 0 \leq t < \tau_1, \\ x_1, & \tau_1 \leq t < \tau_2, \\ x_2, & \tau_2 \leq t < \tau_3, \\ \vdots \end{cases} \tag{1}$$

The process defined by (1) is called a *jump process*. At first glance it might appear that (1) defines $X(t)$ for all $t \geq 0$. But this is not necessarily the case.

Consider, for example, a ball bouncing on the floor. Let the state of the system be the number of bounces it has made. We make the physically reasonable assumption that the time in seconds between the nth bounce and the $(n + 1)$th bounce is 2^{-n}. Then $x_n = n$ and

$$\tau_n = 1 + \frac{1}{2} + \cdots + \frac{1}{2^{n-1}} = 2 - \frac{1}{2^{n-1}}.$$

We see that $\tau_n < 2$ and $\tau_n \to 2$ as $n \to \infty$. Thus (1) defines $X(t)$ only for $0 \le t < 2$. By the time $t = 2$ the ball will have made an infinite number of bounces. In this case it would be appropriate to define $X(t) = \infty$ for $t \ge 2$.

In general, if

$$\lim_{n\to\infty} \tau_n < \infty, \tag{2}$$

we say that the $X(t)$ process *explodes*. If the $X(t)$ process does not explode, i.e., if

$$\lim_{n\to\infty} \tau_n = \infty, \tag{3}$$

then (1) *does* define $X(t)$ for all $t \ge 0$.

We will now specify a probability structure for such a jump process. We suppose that all states are of one of two types, *absorbing* or *non-absorbing*. Once the process reaches an absorbing state, it remains there permanently. With each non-absorbing state x, there is associated a distribution function $F_x(t)$, $-\infty < t < \infty$, which vanishes for $t \le 0$, and *transition probabilities* Q_{xy}, $y \in \mathscr{S}$, which are nonnegative and such that $Q_{xx} = 0$ and

$$\sum_y Q_{xy} = 1. \tag{4}$$

A process starting at x remains there for a random length of time τ_1 having distribution function F_x and then jumps to state $X(\tau_1) = y$ with probability Q_{xy}, $y \in \mathscr{S}$. We assume that τ_1 and $X(\tau_1)$ are chosen independently of each other, i.e., that

$$P_x(\tau_1 \le t, X(\tau_1) = y) = F_x(t)Q_{xy}.$$

Here, as in the previous chapters, we use the notation $P_x(\ )$ and $E_x(\ )$ to denote probabilities of events and expectations of random variables defined in terms of a process initially in state x. Whenever and however the process jumps to a state y, it acts just as a process starting initially at y. For example, if x and y are both non-absorbing states,

$$P_x(\tau_1 \le s, X(\tau_1) = y, \tau_2 - \tau_1 \le t, X(\tau_2) = z) = F_x(s)Q_{xy}F_y(t)Q_{yz}.$$

Similar formulas hold for events defined in terms of three or more jumps. If x is an absorbing state, we set $Q_{xy} = \delta_{xy}$, where

$$\delta_{xy} = \begin{cases} 1, & y = x, \\ 0, & y \ne x. \end{cases}$$

Equation (4) now holds for all $x \in \mathscr{S}$.

We say that the jump process is *pure* or *non-explosive* if (3) holds with probability one regardless of the starting point. Otherwise we say the process is *explosive*. If the state space $\mathscr{S}$ is finite, the jump process is necessarily non-explosive. It is easy to construct examples having an infinite state space which are explosive. Such processes, however, are unlikely to arise in practical applications. At any rate, to keep matters simple we assume that our process is non-explosive. The set of probability zero where (3) fails to hold can safely be ignored. We see from (1) that $X(t)$ is then defined for all $t \geq 0$.

Let $P_{xy}(t)$ denote the probability that a process starting in state x will be in state y at time t. Then

$$P_{xy}(t) = P_x(X(t) = y)$$

and

$$\sum_y P_{xy}(t) = 1.$$

In particular, $P_{xy}(0) = \delta_{xy}$. We can also choose the initial state x according to an *initial distribution* $\pi_0(x)$, $x \in \mathscr{S}$, where $\pi_0(x) \geq 0$ and

$$\sum_x \pi_0(x) = 1.$$

In this case,

$$P(X(t) = y) = \sum_x \pi_0(x)P_{xy}(t).$$

The *transition function* $P_{xy}(t)$ cannot be used directly to obtain such probabilities as

$$P(X(t_1) = x_1, \ldots, X(t_n) = x_n)$$

unless the jump process satisfies the *Markov property*, which states that for $0 \leq s_1 \leq \cdots \leq s_n \leq s \leq t$ and $x_1, \ldots, x_n, x, y \in \mathscr{S}$,

$$P(X(t) = y \mid X(s_1) = x_1, \ldots, X(s_n) = x_n, X(s) = x) = P_{xy}(t - s).$$

By a *Markov pure jump process* we mean a pure jump process that satisfies the Markov property. It can be shown, although not at the level of this book, that a pure jump process is Markovian if and only if all non-absorbing states x are such that

$$P_x(\tau_1 > t + s \mid \tau_1 > s) = P_x(\tau_1 > t), \qquad s, t \geq 0,$$

i.e., such that

$$(5) \qquad \frac{1 - F_x(t + s)}{1 - F_x(s)} = 1 - F_x(t), \qquad s, t \geq 0.$$

Now a distribution function F_x satisfies (5) if and only if it is an exponential distribution function (see Chapter 5 of *Introduction to Probability Theory*). We conclude that a pure jump process is Markovian if and only if F_x is an exponential distribution for all non-absorbing states x.

Let $X(t)$, $0 \le t < \infty$, be a Markov pure jump process. If x is a non-absorbing state, then F_x has an exponential density f_x. Let q_x denote the parameter of this density. Then $q_x = 1/E_x(\tau_1) > 0$ and

$$f_x(t) = \begin{cases} q_x e^{-q_x t}, & t \ge 0, \\ 0, & t < 0. \end{cases}$$

Observe that

$$P_x(\tau_1 \ge t) = \int_t^\infty q_x e^{-q_x s}\, ds = e^{-q_x t}, \qquad t \ge 0.$$

If x is an absorbing state, we set $q_x = 0$.

It follows from the Markov property that for $0 \le t_1 \le \cdots \le t_n$ and $x_1, \ldots, x_n$ in $\mathcal{S}$,

$$\text{(6)} \qquad \begin{aligned} &P(X(t_1) = x_1, \ldots, X(t_n) = x_n) \\ &\qquad = P(X(t_1) = x_1)P_{x_1x_2}(t_2 - t_1)\cdots P_{x_{n-1}x_n}(t_n - t_{n-1}). \end{aligned}$$

In particular, for $s \ge 0$ and $t \ge 0$

$$P_x(X(t) = z, X(t + s) = y) = P_{xz}(t)P_{zy}(s).$$

Since

$$P_{xy}(t + s) = \sum_z P_x(X(t) = z, X(t + s) = y),$$

we conclude that

$$\text{(7)} \qquad P_{xy}(t + s) = \sum_z P_{xz}(t)P_{zy}(s), \qquad s \ge 0 \text{ and } t \ge 0.$$

Equation (7) is known as the *Chapman-Kolmogorov* equation.

The transition function $P_{xy}(t)$ satisfies the integral equation

$$\text{(8)} \qquad P_{xy}(t) = \delta_{xy}e^{-q_x t} + \int_0^t q_x e^{-q_x s}\left(\sum_{z \ne x} Q_{xz}P_{zy}(t - s)\right) ds, \qquad t \ge 0,$$

which we will now verify. If x is an absorbing state, (8) reduces to the obvious fact that

$$P_{xy}(t) = \delta_{xy}, \qquad t \ge 0.$$

Suppose x is not an absorbing state. Then for a process starting at x, the event $\{\tau_1 \le t, X(\tau_1) = z \text{ and } X(t) = y\}$ occurs if and only if the first jump occurs at some time $s \le t$ and takes the process to z, and the process goes from z to y in the remaining $t - s$ units of time. Thus

$$P_x(\tau_1 \le t, X(\tau_1) = z \text{ and } X(t) = y) = \int_0^t q_x e^{-q_x s} Q_{xz}P_{zy}(t - s)\, ds,$$

so

$$P_x(\tau_1 \le t \text{ and } X(t) = y) = \sum_{z \ne x} P_x(\tau_1 \le t,\ X(\tau_1) = z \text{ and } X(t) = y)$$
$$= \int_0^t q_x e^{-q_x s} \left(\sum_{z \ne x} Q_{xz} P_{zy}(t - s) \right) ds.$$

Also

$$P_x(\tau_1 > t \text{ and } X(t) = y) = \delta_{xy} P_x(\tau_1 > t)$$
$$= \delta_{xy} e^{-q_x t}.$$

Consequently,

$$P_{xy}(t) = P_x(X(t) = y)$$
$$= P_x(\tau_1 > t \text{ and } X(t) = y) + P_x(\tau_1 \le t \text{ and } X(t) = y)$$
$$= \delta_{xy} e^{-q_x t} + \int_0^t q_x e^{-q_x s} \left(\sum_{z \ne x} Q_{xz} P_{zy}(t - s) \right) ds,$$

as claimed. Replacing s by $t - s$ in the integral in (8), we can rewrite (8) as

$$P_{xy}(t) = \delta_{xy} e^{-q_x t} + q_x e^{-q_x t} \int_0^t e^{q_x s} \left(\sum_{z \ne x} Q_{xz} P_{zy}(s) \right) ds, \qquad t \ge 0. \tag{9}$$

It follows from (9) that $P_{xy}(t)$ is continuous in t for $t \ge 0$. Therefore the integrand in (9) is a continuous function, so we can differentiate the right side. We obtain

$$P'_{xy}(t) = -q_x P_{xy}(t) + q_x \sum_{z \ne x} Q_{xz} P_{zy}(t), \qquad t \ge 0. \tag{10}$$

In particular,

$$P'_{xy}(0) = -q_x P_{xy}(0) + q_x \sum_{z \ne x} Q_{xz} P_{zy}(0)$$
$$= -q_x \delta_{xy} + q_x \sum_{z \ne x} Q_{xz} \delta_{zy}$$
$$= -q_x \delta_{xy} + q_x Q_{xy}.$$

Set

$$q_{xy} = P'_{xy}(0), \qquad x, y \in \mathscr{S}. \tag{11}$$

Then

$$q_{xy} = \begin{cases} -q_x, & y = x, \\ q_x Q_{xy}, & y \ne x. \end{cases} \tag{12}$$

It follows from (12) that

$$\sum_{y \ne x} q_{xy} = q_x = -q_{xx}. \tag{13}$$

The quantities q_{xy}, $x \in \mathscr{S}$ and $y \in \mathscr{S}$, are called the *infinitesimal parameters* of the process. These parameters determine q_x and Q_{xy}, and thus by our construction determine a unique Markov pure jump process. We can rewrite (10) in terms of the infinitesimal parameters as

$$(14) \qquad P'_{xy}(t) = \sum_z q_{xz} P_{zy}(t), \qquad t \geq 0.$$

This equation is known as the *backward equation.*

If $\mathscr{S}$ is finite, we can differentiate the Chapman-Kolmogorov equation with respect to s, obtaining

$$(15) \qquad P'_{xy}(t + s) = \sum_z P_{xz}(t) P'_{zy}(s), \qquad s \geq 0 \text{ and } t \geq 0.$$

In particular,

$$P'_{xy}(t) = \sum_z P_{xz}(t) P'_{zy}(0), \qquad t \geq 0,$$

or equivalently,

$$(16) \qquad P'_{xy}(t) = \sum_z P_{xz}(t) q_{zy}, \qquad t \geq 0.$$

Formula (16) is known as the *forward equation.* It can be shown that (15) and (16) hold even if $\mathscr{S}$ is infinite, but the proofs are not easy and will be omitted.

In Section 3.2 we will describe some examples in which the backward or forward equation can be used to find explicit formulas for $P_{xy}(t)$.

3.2. Birth and death processes

Let $\mathscr{S} = \{0, 1, \ldots, d\}$ or $\mathscr{S} = \{0, 1, 2, \ldots\}$. By a *birth and death process* on $\mathscr{S}$ we mean a Markov pure jump process on $\mathscr{S}$ having infinitesimal parameters q_{xy} such that

$$q_{xy} = 0, \qquad |y - x| > 1.$$

Thus a birth and death process starting at x can in one jump go only to the states $x - 1$ or $x + 1$.

The parameters $\lambda_x = q_{x,x+1}$, $x \in \mathscr{S}$, and $\mu_x = q_{x,x-1}$, $x \in \mathscr{S}$, are called respectively the *birth rates* and *death rates* of the process. The parameters q_x and Q_{xy} of the process can be expressed simply in terms of the birth and death rates. By (13)

$$-q_{xx} = q_x = q_{x,x+1} + q_{x,x-1},$$

so that

$$(17) \qquad q_{xx} = -(\lambda_x + \mu_x) \qquad \text{and} \qquad q_x = \lambda_x + \mu_x.$$

Thus x is an absorbing state if and only if $\lambda_x = \mu_x = 0$. If x is a non-absorbing state, then by (12)

$$(18) \qquad Q_{xy} = \begin{cases} \dfrac{\mu_x}{\lambda_x + \mu_x}, & y = x - 1, \\ \dfrac{\lambda_x}{\lambda_x + \mu_x}, & y = x + 1, \\ 0, & \text{elsewhere.} \end{cases}$$

A birth and death process is called a *pure birth process* if $\mu_x = 0$, $x \in \mathscr{S}$, and a *pure death process* if $\lambda_x = 0$, $x \in \mathscr{S}$. A pure birth process can move only to the right, and a pure death process can move only to the left.

Given nonnegative numbers λ_x, $x \in \mathscr{S}$, and μ_x, $x \in \mathscr{S}$, it is natural to ask whether there is a birth and death process corresponding to these parameters. Of course, $\mu_0 = 0$ is a necessary requirement, as is $\lambda_d = 0$ if $\mathscr{S}$ is finite. The only additional problem is that explosions must be ruled out if $\mathscr{S}$ is infinite. It is not difficult to derive a necessary and sufficient condition for the process to be non-explosive. A simple sufficient condition for the process to be non-explosive is that for some positive numbers A and B

$$\lambda_x \le A + Bx, \qquad x \ge 0.$$

This condition holds in all the examples we will consider.

In finding the birth and death rates of specific processes, we will use some standard properties of independent exponentially distributed random variables. Let $\xi_1, \ldots, \xi_n$ be independent random variables having exponential distributions with respective parameters $\alpha_1, \ldots, \alpha_n$. Then $\min(\xi_1, \ldots, \xi_n)$ has an exponential distribution with parameter $\alpha_1 + \cdots + \alpha_n$ and

$$(19) \qquad P(\xi_k = \min(\xi_1, \ldots, \xi_n)) = \frac{\alpha_k}{\alpha_1 + \cdots + \alpha_n}, \qquad k = 1, \ldots, n.$$

Moreover, with probability one, the random variables $\xi_1, \ldots, \xi_n$ take on n distinct values.

To verify these results we observe first that

$$\begin{aligned} P(\min(\xi_1, \ldots, \xi_n) > t) &= P(\xi_1 > t, \ldots, \xi_n > t) \\ &= P(\xi_1 > t) \cdots P(\xi_n > t) \\ &= e^{-\alpha_1 t} \cdots e^{-\alpha_n t} \\ &= e^{-(\alpha_1 + \cdots + \alpha_n)t}, \end{aligned}$$

and hence that $\min(\xi_1, \ldots, \xi_n)$ has the indicated exponential distribution.

Set

$$\eta_k = \min(\xi_j : j \neq k).$$

Then η_k has an exponential distribution with parameter

$$\beta_k = \sum_{j \neq k} \alpha_j,$$

and ξ_k and η_k are independent. Thus

$$\begin{aligned} P(\xi_k = \min(\xi_1, \ldots, \xi_n)) &= P(\xi_k \leq \eta_k) \\ &= \int_0^\infty \left(\int_x^\infty \alpha_k e^{-\alpha_k x} \beta_k e^{-\beta_k y} \, dy \right) dx \\ &= \int_0^\infty \alpha_k e^{-\alpha_k x} e^{-\beta_k x} \, dx \\ &= \frac{\alpha_k}{\alpha_k + \beta_k} = \frac{\alpha_k}{\alpha_1 + \cdots + \alpha_n}. \end{aligned}$$

In order to show that the random variables $\xi_1, \ldots, \xi_n$ take on n distinct values with probability one, it is enough to show that $P(\xi_i \neq \xi_j) = 1$ for $i \neq j$. But since ξ_i and ξ_j have a joint density f, it follows that

$$P(\xi_i = \xi_j) = \iint_{\{(x, y): x = y\}} f(x, y) \, dx \, dy = 0,$$

as desired.

Example 1. Branching process. Consider a collection of particles which act independently in giving rise to succeeding generations of particles. Suppose that each particle, from the time it appears, waits a random length of time having an exponential distribution with parameter q and then splits into two identical particles with probability p and disappears with probability $1 - p$. Let $X(t)$, $0 \leq t < \infty$, denote the number of particles present at time t. This branching process is a birth and death process. Find the birth and death rates.

Consider a branching process starting out with x particles. Let $\xi_1, \ldots, \xi_x$ be the times until these particles split apart or disappear. Then $\xi_1, \ldots, \xi_x$ each has an exponential distribution with parameter q, and hence $\tau_1 = \min(\xi_1, \ldots, \xi_x)$ has an exponential distribution with parameter $q_x = xq$. Whichever particle acts first has probability p of splitting into two particles and probability $1 - p$ of disappearing. Thus for $x \geq 1$

$$Q_{x,x+1} = p \qquad \text{and} \qquad Q_{x,x-1} = 1 - p.$$

State 0 is an absorbing state. Since $\lambda_x = q_x Q_{x,x+1}$ and $\mu_x = q_x Q_{x,x-1}$, we conclude that

$$\lambda_x = xqp \qquad \text{and} \qquad \mu_x = xq(1 - p), \qquad x \geq 0.$$

In the preceding example we did not actually prove that the process is a birth and death process, i.e., that it "starts from scratch" after making a jump. This intuitively reasonable property basically depends on the fact that an exponentially distributed random variable ξ satisfies the formula

$$P(\xi > t + s \mid \xi > s) = P(\xi > t), \qquad s, t \geq 0,$$

but a rigorous proof is complicated.

By (17) and the definition of λ_x and μ_x, the backward and forward equations for a birth and death process can be written respectively as

$$(20) \qquad P'_{xy}(t) = \mu_x P_{x-1,y}(t) - (\lambda_x + \mu_x)P_{xy}(t) + \lambda_x P_{x+1,y}(t), \qquad t \geq 0,$$

and

$$(21) \qquad P'_{xy}(t) = \lambda_{y-1} P_{x,y-1}(t) - (\lambda_y + \mu_y)P_{xy}(t) + \mu_{y+1} P_{x,y+1}(t), \qquad t \geq 0.$$

In (21) we set $\lambda_{-1} = 0$, and if $\mathscr{S} = \{0, \ldots, d\}$ for $d < \infty$, we set $\mu_{d+1} = 0$.

We will solve the backward and forward equations for a birth and death process in some special cases. To do so we will use the result that if

$$(22) \qquad f'(t) = -\alpha f(t) + g(t), \qquad t \geq 0,$$

then

$$(23) \qquad f(t) = f(0)e^{-\alpha t} + \int_0^t e^{-\alpha(t-s)} g(s)\, ds, \qquad t \geq 0.$$

The proof of this standard result is very easy. We multiply (22) through by $e^{\alpha t}$ and rewrite the resulting equation as

$$\frac{d}{dt}\left(e^{\alpha t} f(t)\right) = e^{\alpha t} g(t).$$

Integrating from 0 to t we find that

$$e^{\alpha t} f(t) - f(0) = \int_0^t e^{\alpha s} g(s)\, ds,$$

and hence that (23) holds.

3.2.1. Two-state birth and death process. Consider a birth and death process having state space $\mathscr{S} = \{0, 1\}$, and suppose that 0 and 1 are both non-absorbing states. Since $\mu_0 = \lambda_1 = 0$, the process is determined by the parameters λ_0 and μ_1. For simplicity in notation we set $\lambda = \lambda_0$ and $\mu = \mu_1$. We can interpret such a process by thinking of state 1 as the system (e.g., telephone or machine) operating and state 0 as the system being idle. We suppose that starting from an idle state the system remains idle for a random length of time which is exponentially distributed with parameter λ, and that starting in an operating state the system continues operating for a random length of time which is exponentially distributed with parameter μ.

We will find the transition function of the process by solving the backward equation. It is left as an exercise for the reader to obtain the same results by solving the forward equation.

Setting $y = 0$ in (20), we see that

$$P'_{00}(t) = -\lambda P_{00}(t) + \lambda P_{10}(t), \qquad t \geq 0, \tag{24}$$

and

$$P'_{10}(t) = \mu P_{00}(t) - \mu P_{10}(t), \qquad t \geq 0. \tag{25}$$

Subtracting the second equation from the first,

$$\frac{d}{dt}(P_{00}(t) - P_{10}(t)) = -(\lambda + \mu)(P_{00}(t) - P_{10}(t)).$$

Applying (23),

$$\begin{aligned} P_{00}(t) - P_{10}(t) &= (P_{00}(0) - P_{10}(0))e^{-(\lambda+\mu)t} \\ &= e^{-(\lambda+\mu)t}. \end{aligned} \tag{26}$$

Here we have used the formulas $P_{00}(0) = 1$ and $P_{10}(0) = 0$. It now follows from (24) that

$$\begin{aligned} P'_{00}(t) &= -\lambda(P_{00}(t) - P_{10}(t)) \\ &= -\lambda e^{-(\lambda+\mu)t}. \end{aligned}$$

Thus

$$\begin{aligned} P_{00}(t) &= P_{00}(0) + \int_0^t P'_{00}(s)\, ds \\ &= 1 - \int_0^t \lambda e^{-(\lambda+\mu)s}\, ds \\ &= 1 - \frac{\lambda}{\lambda + \mu}(1 - e^{-(\lambda+\mu)t}), \end{aligned}$$

or equivalently,

$$P_{00}(t) = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} e^{-(\lambda+\mu)t}, \qquad t \geq 0. \tag{27}$$

Now, by (26), $P_{10}(t) = P_{00}(t) - e^{-(\lambda+\mu)t}$, and therefore

$$(28) \qquad P_{10}(t) = \frac{\mu}{\lambda + \mu} - \frac{\mu}{\lambda + \mu} e^{-(\lambda+\mu)t}, \qquad t \geq 0.$$

By setting $y = 1$ in the backward equation, or by subtracting $P_{00}(t)$ and $P_{10}(t)$ from one, we conclude that

$$(29) \qquad P_{01}(t) = \frac{\lambda}{\lambda + \mu} - \frac{\lambda}{\lambda + \mu} e^{-(\lambda+\mu)t}, \qquad t \geq 0,$$

and

$$(30) \qquad P_{11}(t) = \frac{\lambda}{\lambda + \mu} + \frac{\mu}{\lambda + \mu} e^{-(\lambda+\mu)t}, \qquad t \geq 0.$$

From (27)–(30) we see that

$$(31) \qquad \lim_{t \to +\infty} P_{xy}(t) = \pi(y),$$

where

$$(32) \qquad \pi(0) = \frac{\mu}{\lambda + \mu} \quad \text{and} \quad \pi(1) = \frac{\lambda}{\lambda + \mu}.$$

If π_0 is the initial distribution of the process, then by (27) and (28)

$$\begin{aligned} P(X(t) = 0) &= \pi_0(0)P_{00}(t) + (1 - \pi_0(0))P_{10}(t) \\ &= \frac{\mu}{\lambda + \mu} + \left(\pi_0(0) - \frac{\mu}{\lambda + \mu}\right) e^{-(\lambda+\mu)t}, \qquad t \geq 0. \end{aligned}$$

Similarly,

$$P(X(t) = 1) = \frac{\lambda}{\lambda + \mu} + \left(\pi_0(1) - \frac{\lambda}{\lambda + \mu}\right) e^{-(\lambda+\mu)t}, \qquad t \geq 0.$$

Thus $P(X(t) = 0)$ and $P(X(t) = 1)$ are independent of t if and only if π_0 is the distribution π given by (32).

3.2.2. Poisson process. Consider a pure birth process $X(t)$, $0 \leq t < \infty$, on the nonnegative integers such that

$$\lambda_x = \lambda > 0, \qquad x \geq 0.$$

Since a pure birth process can move only to the right,

$$(33) \qquad P_{xy}(t) = 0, \qquad y < x \text{ and } t \geq 0.$$

Also $P_{xx}(t) = P_x(\tau_1 > t)$ and hence

$$(34) \qquad P_{xx}(t) = e^{-\lambda t}, \qquad t \geq 0.$$

The forward equation for $y \neq 0$ is

$$P'_{xy}(t) = \lambda P_{x,y-1}(t) - \lambda P_{xy}(t), \qquad t \geq 0.$$

From (23) we see that

$$P_{xy}(t) = e^{-\lambda t} P_{xy}(0) + \lambda \int_0^t e^{-\lambda(t-s)} P_{x,y-1}(s)\, ds, \qquad t \geq 0.$$

Since $P_{xy}(0) = \delta_{xy}$, we conclude that for $y > x$

(35) $$P_{xy}(t) = \lambda \int_0^t e^{-\lambda(t-s)} P_{x,y-1}(s)\, ds, \qquad t \geq 0.$$

It follows from (34) and (35) that

$$P_{x,x+1}(t) = \lambda \int_0^t e^{-\lambda(t-s)} e^{-\lambda s}\, ds = \lambda e^{-\lambda t} \int_0^t ds = \lambda t e^{-\lambda t}$$

and hence by using (35) once more that

$$P_{x,x+2}(t) = \lambda \int_0^t e^{-\lambda(t-s)} \lambda s e^{-\lambda s}\, ds = \lambda^2 e^{-\lambda t} \int_0^t s\, ds = \frac{(\lambda t)^2}{2} e^{-\lambda t}.$$

By induction

(36) $$P_{xy}(t) = \frac{(\lambda t)^{y-x} e^{-\lambda t}}{(y-x)!}, \qquad 0 \leq x \leq y \text{ and } t \geq 0.$$

Formulas (33) and (36) imply that

(37) $$P_{xy}(t) = P_{0,y-x}(t), \qquad t \geq 0,$$

and that if $X(0) = x$, then $X(t) - x$ has a Poisson distribution with parameter λt.

In general, for $0 \leq s \leq t$, $X(t) - X(s)$ has a Poisson distribution with parameter $\lambda(t - s)$. For if $0 \leq s \leq t$ and y is a nonnegative integer, then

$$\begin{aligned} P(X(t) - X(s) = y) &= \sum_x P(X(s) = x \text{ and } X(t) = x + y) \\ &= \sum_x P(X(s) = x) P_{x,x+y}(t - s) \\ &= \sum_x P(X(s) = x) P_{0y}(t - s) \\ &= P_{0y}(t - s) \\ &= \frac{(\lambda(t - s))^y e^{-\lambda(t-s)}}{y!}. \end{aligned}$$

If $0 \leq t_1 \leq \cdots \leq t_n$, the random variables

$$X(t_2) - X(t_1), \ldots, X(t_n) - X(t_{n-1})$$

are independent. For we observe that if $z_1, \ldots, z_{n-1}$ are arbitrary integers, then by (6) and (37)

$$
\begin{aligned}
P(X(t_2) - X(t_1) = z_1, \ldots, X(t_n) - X(t_{n-1}) = z_{n-1}) \\
&= \sum_x P(X(t_1) = x) P_{0z_1}(t_2 - t_1) \cdots P_{0z_{n-1}}(t_n - t_{n-1}) \\
&= P_{0z_1}(t_2 - t_1) \cdots P_{0z_{n-1}}(t_n - t_{n-1}) \\
&= P(X(t_2) - X(t_1) = z_1) \cdots P(X(t_n) - X(t_{n-1}) = z_{n-1}).
\end{aligned}
$$

By a *Poisson process with parameter* λ on $0 \le t < \infty$, we mean a pure birth process $X(t)$, $0 \le t < \infty$, having state space $\{0, 1, 2, \ldots\}$, constant birth rate $\lambda_x = \lambda > 0$, and initial value $X(0) = 0$. According to the above discussion the Poisson process satisfies the following three properties:

(i) $X(0) = 0$.
(ii) $X(t) - X(s)$ has a Poisson distribution with parameter $\lambda(t - s)$ for $0 \le s \le t$.
(iii) $X(t_2) - X(t_1), X(t_3) - X(t_2), \ldots, X(t_n) - X(t_{n-1})$ are independent for $0 \le t_1 \le t_2 \le \cdots \le t_n$.

The Poisson process can be used to model events occurring in time, such as calls coming into a telephone exchange, customers arriving at a queue, and radioactive disintegrations. Let $X(t)$, $0 \le t < \infty$, denote the number of events occurring in the time interval $(0, t]$. For $0 \le s \le t$ the random variable $X(t) - X(s)$ denotes the number of events in the time interval $(s, t]$. If the waiting times between successive events are independent and exponentially distributed with common parameter λ, then $X(t)$, $0 \le t < \infty$, is a Poisson process and properties (i)–(iii) hold. Property (ii) states that the number of events in any interval has a Poisson distribution. Property (iii) states that the numbers of events in disjoint time intervals are independent. Conversely, if $X(t)$, $0 \le t < \infty$, satisfies properties (i)–(iii), then the waiting times between successive events are independent and exponentially distributed with common parameter λ, and hence $X(t)$ is a pure birth process with constant birth rate λ. This result was proved in Chapter 9 of Volume I, but will not be needed.

Since the Poisson process is a pure birth process starting in state 0, it follows that for $n \ge 1$ the time τ_n of the nth jump equals the time T_n when the process hits state n. When the Poisson process is used to model events occurring in time as described above, the common time $\tau_n = T_n$ is the time when the nth event occurs.

The Poisson process can be used to construct a variety of other processes.

Example 2. Branching process with immigration. Consider the branching process introduced in Example 1. Suppose that new particles immigrate into the system at random times that form a Poisson process with parameter λ and then give rise to succeeding generations as described in Example 1. Find the birth and death rates of this birth and death process.

Suppose there are initially x particles present. Let $\xi_1, \ldots, \xi_x$ be the times at which these particles split apart or disappear, and let η be the first time a new particle enters the system. We interpret the description of the system as implying that η is independent of $\xi_1, \ldots, \xi_x$. Then $\xi_1, \ldots, \xi_x$, η are independent exponentially distributed random variables having respective parameters $q, \ldots, q, \lambda$. Thus

$$\tau_1 = \min(\xi_1, \ldots, \xi_x, \eta)$$

is exponentially distributed with parameter $q_x = xq + \lambda$, and by (19)

$$P(\tau_1 = \eta) = \frac{\lambda}{xq + \lambda}.$$

The event $\{X(\tau_1) = x + 1\}$ occurs if either $\tau_1 = \eta$ or

$$\tau_1 = \min(\xi_1, \ldots, \xi_x)$$

and a particle splits into two new particles at time τ_1. Thus

$$Q_{x,x+1} = \frac{\lambda}{xq + \lambda} + \frac{xq}{xq + \lambda} p.$$

Also,

$$Q_{x,x-1} = \frac{xq}{xq + \lambda} (1 - p).$$

We conclude that

$$\lambda_x = q_x Q_{x,x+1} = xqp + \lambda$$

and

$$\mu_x = q_x Q_{x,x-1} = xq(1 - p).$$

It is also possible to construct a Poisson process with parameter λ on $-\infty < t < \infty$. We first construct two independent Poisson processes $X_1(t)$, $0 \le t < \infty$, and $X_2(t)$, $0 \le t < \infty$, both having parameter λ. We then define $X(t)$, $-\infty < t < \infty$, by

$$X(t) = \begin{cases} -X_1(-t), & t < 0, \\ X_2(t), & t \ge 0. \end{cases}$$

It is easy to show that the process $X(t)$, $-\infty < t < \infty$, so constructed, satisfies the following three properties:

(i) $X(0) = 0$.
(ii) $X(t) - X(s)$ has a Poisson distribution with parameter $\lambda(t - s)$ for $s \le t$.
(iii) $X(t_2) - X(t_1), \ldots, X(t_n) - X(t_{n-1})$ are independent for $t_1 \le t_2 \le \cdots \le t_n$.

3.2.3. Pure birth process. Consider a pure birth process $X(t)$, $0 \le t < \infty$, on $\{0, 1, 2, \ldots\}$. The forward equation (21) reduces to

$$(38) \qquad P'_{xy}(t) = \lambda_{y-1}P_{x,y-1}(t) - \lambda_y P_{xy}(t), \qquad t \ge 0.$$

Since the process moves only to the right,

$$(39) \qquad P_{xy}(t) = 0, \qquad y < x \text{ and } t \ge 0.$$

It follows from (38) and (39) that

$$P'_{xx}(t) = -\lambda_x P_{xx}(t).$$

Since $P_{xx}(0) = 1$ and $P_{xy}(0) = 0$ for $y > x$, we conclude from (23) that

$$(40) \qquad P_{xx}(t) = e^{-\lambda_x t}, \qquad t \ge 0,$$

and

$$(41) \qquad P_{xy}(t) = \lambda_{y-1}\int_0^t e^{-\lambda_y(t-s)}P_{x,y-1}(s)\,ds, \qquad y > x \text{ and } t \ge 0.$$

We can use (40) and (41) to find $P_{xy}(t)$ recursively for $y > x$. In particular,

$$P_{x,x+1}(t) = \lambda_x \int_0^t e^{-\lambda_{x+1}(t-s)}e^{-\lambda_x s}\,ds,$$

and hence for $t \ge 0$

$$(42) \qquad P_{x,x+1}(t) = \begin{cases} \dfrac{\lambda_x}{\lambda_{x+1} - \lambda_x}(e^{-\lambda_x t} - e^{-\lambda_{x+1} t}), & \lambda_{x+1} \ne \lambda_x, \\ \lambda_x t e^{-\lambda_x t}, & \lambda_{x+1} = \lambda_x. \end{cases}$$

Example 3. Linear birth process. Consider a pure birth process on $\{0, 1, 2, \ldots\}$ having birth rates

$$\lambda_x = x\lambda, \qquad x \ge 0,$$

for some positive constant λ (the branching process with $p = 1$ is of this form). Find $P_{xy}(t)$.

As noted above, $P_{xy}(t) = 0$ for $y < x$ and

$$P_{xx}(t) = e^{-\lambda_x t} = e^{-x\lambda t}.$$

We see from (42) that

$$P_{x,x+1}(t) = xe^{-x\lambda t}(1 - e^{-\lambda t}).$$

To compute $P_{x,x+2}(t)$ we set $y = x + 2$ in (41) and obtain

$$\begin{aligned} P_{x,x+2}(t) &= (x + 1)x\lambda \int_0^t e^{-(x+2)\lambda(t-s)}e^{-x\lambda s}(1 - e^{-\lambda s})\, ds \\ &= (x + 1)x\lambda e^{-(x+2)\lambda t} \int_0^t e^{2\lambda s}(1 - e^{-\lambda s})\, ds \\ &= (x + 1)x\lambda e^{-(x+2)\lambda t} \int_0^t e^{\lambda s}(e^{\lambda s} - 1)\, ds \\ &= (x + 1)x\lambda e^{-(x+2)\lambda t} \frac{(e^{\lambda t} - 1)^2}{2\lambda} \\ &= \binom{x+1}{2} e^{-x\lambda t}(1 - e^{-\lambda t})^2. \end{aligned}$$

It is left as an exercise for the reader to show by induction that

$$(43) \qquad P_{xy}(t) = \binom{y-1}{y-x} e^{-x\lambda t}(1 - e^{-\lambda t})^{y-x}, \qquad y \geq x \text{ and } t \geq 0.$$

3.2.4. Infinite server queue. Suppose that customers arrive for service according to a Poisson process with parameter λ and that each customer starts being served immediately upon his arrival (i.e., that there are an infinite number of servers). Suppose that the service times are independent and exponentially distributed with parameter μ. Let $X(t)$, $0 \leq t < \infty$, denote the number of customers in the process of being served at time t. This birth and death process, called an *infinite server queue*, is a special case of the branching process with immigration corresponding to $q = \mu$ and $p = 0$. We conclude that $\lambda_x = \lambda$ and $\mu_x = x\mu$, $x \geq 0$. The transition function $P_{xy}(t)$ will now be obtained by a probabilistic argument.

Let $Y(t)$ denote the number of customers who arrive in the time interval $(0, t]$. An interesting and useful result about the Poisson process is that conditioned on $Y(t) = k$, the times when the first k customers arrive are distributed as k independent random variables each uniformly distributed on $(0, t]$. In order to see intuitively why this should be true, consider an arbitrary partition $0 = t_0 < t_1 < \cdots < t_m = t$ of $[0, t]$ and let X_i denote the number of customers arriving between time t_{i-1} and time t_i. Then $X_1, \ldots, X_m$ are independent random variables having Poisson distributions with respective parameters

$$\lambda(t_1 - t_0), \ldots, \lambda(t_m - t_{m-1}),$$

and $X_1 + \cdots + X_m = Y(t)$ has a Poisson distribution with parameter λt. Thus for $x_1, \ldots, x_m$ nonnegative integers adding up to k,

$$\begin{aligned}
P(X_1 = x_1, \ldots, X_m = x_m \mid Y(t) = k) \\
&= P(X_1 = x_1, \ldots, X_m = x_m \mid X_1 + \cdots + X_m = k) \\
&= \frac{P(X_1 = x_1, \ldots, X_m = x_m, X_1 + \cdots + X_m = k)}{P(X_1 + \cdots + X_m = k)} \\
&= \frac{P(X_1 = x_1, \ldots, X_m = x_m)}{P(X_1 + \cdots + X_m = k)} \\
&= \frac{\displaystyle\prod_{i=1}^{m} \frac{[\lambda(t_i - t_{i-1})]^{x_i} e^{-\lambda(t_i - t_{i-1})}}{x_i!}}{\dfrac{(\lambda t)^k e^{-\lambda t}}{k!}} \\
&= \frac{k!}{\prod_{i=1}^{m} x_i!} \prod_{i=1}^{m} \left(\frac{t_i - t_{i-1}}{t}\right)^{x_i}.
\end{aligned}$$

But these multinomial probabilities are just those that would be obtained by choosing the k arrival times independently and uniformly distributed over $(0, t]$.

If a customer arrives at time $s \in (0, t]$, the probability that he is still in the process of being served at time t is $e^{-\mu(t-s)}$. Thus if a customer arrives at a time chosen uniformly from $(0, t]$, the probability that he is still in the process of being served at time t is

$$p_t = \frac{1}{t} \int_0^t e^{-\mu(t-s)}\, ds = \frac{1 - e^{-\mu t}}{\mu t}.$$

Let $X_1(t)$ denote the number of customers arriving in $(0, t]$ that are still in the process of being served at time t. It follows from the results of the previous two paragraphs that the conditional distribution of $X_1(t)$ given that $Y(t) = k$ is a binomial distribution with parameters k and p_t, i.e., that

$$P(X_1(t) = n \mid Y(t) = k) = \binom{k}{n} p_t^n (1 - p_t)^{k-n}.$$

Since $Y(t)$ has a Poisson distribution with parameter λt, we conclude that

$$
\begin{aligned}
P(X_1(t) = n) &= \sum_{k=n}^{\infty} P(Y(t) = k,\ X_1(t) = n) \\
&= \sum_{k=n}^{\infty} P(Y(t) = k)P(X_1(t) = n \mid Y(t) = k) \\
&= \sum_{k=n}^{\infty} \frac{(\lambda t)^k e^{-\lambda t}}{k!} \frac{k!}{n!\,(k-n)!} p_t^n (1 - p_t)^{k-n} \\
&= \frac{(\lambda t p_t)^n e^{-\lambda t}}{n!} \sum_{k=n}^{\infty} \frac{(\lambda t(1 - p_t))^{k-n}}{(k-n)!} \\
&= \frac{(\lambda t p_t)^n e^{-\lambda t}}{n!} \sum_{m=0}^{\infty} \frac{(\lambda t(1 - p_t))^m}{m!} \\
&= \frac{(\lambda t p_t)^n e^{-\lambda t}}{n!} e^{\lambda t(1 - p_t)} \\
&= \frac{(\lambda t p_t)^n e^{-\lambda t p_t}}{n!}.
\end{aligned}
$$

Thus $X_1(t)$ has a Poisson distribution with parameter

$$\lambda t p_t = \frac{\lambda}{\mu}(1 - e^{-\mu t}).$$

Let x denote the number of customers present initially and let $X_2(t)$ denote the number of these customers still in the process of being served at time t. Then $X_2(t)$ is independent of $X_1(t)$ and has a binomial distribution with parameters x and $e^{-\mu t}$. Since $X(t) = X_1(t) + X_2(t)$, we conclude that

$$P_{xy}(t) = P_x(X(t) = y) = \sum_{k=0}^{\min(x,y)} P_x(X_2(t) = k)P(X_1(t) = y - k).$$

Therefore

$$
\begin{aligned}
(44)\qquad P_{xy}(t) = \sum_{k=0}^{\min(x,y)} \Bigg[&\binom{x}{k} e^{-k\mu t}(1 - e^{-\mu t})^{x-k} \\
&\times \frac{\left(\frac{\lambda}{\mu}(1 - e^{-\mu t})\right)^{y-k}}{(y-k)!} \exp\left(-\frac{\lambda}{\mu}(1 - e^{-\mu t})\right)\Bigg].
\end{aligned}
$$

As $t \to \infty$, $e^{-\mu t} \to 0$, and hence the terms in the sum in (44) all approach 0 except the term corresponding to $k = 0$. Consequently

$$(45)\qquad \lim_{t\to\infty} P_{xy}(t) = \frac{(\lambda/\mu)^y e^{-\lambda/\mu}}{y!}.$$

3.3. Properties of a Markov pure jump process

In this section we will discuss the notions of recurrence, transience, irreducibility, stationary distributions, and positive recurrence of Markov pure jump processes. The results will be described briefly and without proofs, as they are very similar to those for the Markov chains discussed in Chapters 1 and 2. In Section 3.3.1 we apply these results to birth and death processes.

Let $X(t)$, $0 \le t < \infty$, be a Markov pure jump process having state space $\mathscr{S}$. For $y \in \mathscr{S}$ and $X(0) \neq y$, the first visit to y takes place at time

$$T_y = \min\,(t \ge 0 : X(t) = y).$$

If $X(0) = y$, then $\min\,(t \ge 0 : X(t) = y) = 0$. A more useful random variable in this case is the time T_y of the first return to y after the process leaves y. Both cases are covered by setting

$$T_y = \min\,(t \ge \tau_1 : X(t) = y).$$

Here τ_1 is the time of the first jump. If $\tau_1 = \infty$ or $X(t) \neq y$ for all $t \ge \tau_1$, we set $T_y = \infty$.

If x is an absorbing state, set $\rho_{xy} = \delta_{xy}$; and if x is a non-absorbing state, set

$$\rho_{xy} = P_x(T_y < \infty).$$

A state $y \in \mathscr{S}$ is called *recurrent* if $\rho_{yy} = 1$ and *transient* if $\rho_{yy} < 1$. The process is said to be a *recurrent process* if all of its states are recurrent and a *transient process* if all of its states are transient. The process is called *irreducible* if $\rho_{xy} > 0$ for all choices of $x \in \mathscr{S}$ and $y \in \mathscr{S}$.

The function $P(x, y) = Q_{xy}$, $x \in \mathscr{S}$ and $y \in \mathscr{S}$, is the transition function of a Markov chain called the *embedded chain.* The quantities ρ_{xy} for this Markov chain are equal to the corresponding quantities for the Markov pure jump process. This relationship shows that results of Chapter 1 involving only the numbers ρ_{xy} are also valid in the present context. In particular, an irreducible process is either a recurrent process or a transient process. It is recurrent if and only if the embedded chain is recurrent.

If $\pi(x)$, $x \in \mathscr{S}$, are nonnegative numbers summing to one and if

$$\sum_x \pi(x)P_{xy}(t) = \pi(y), \qquad y \in \mathscr{S} \text{ and } t \ge 0, \tag{46}$$

then π is called a *stationary distribution.* If $X(0)$ has a stationary distribution π for its initial distribution, then

$$P(X(t) = y) = \sum_x \pi(x)P_{xy}(t) = \pi(y),$$

so that $X(t)$ has distribution π for all $t \ge 0$.

If (46) holds and $\mathscr{S}$ is finite, we can differentiate this equation and obtain

$$\sum_x \pi(x)P'_{xy}(t) = 0, \qquad y \in \mathscr{S} \text{ and } t \geq 0. \tag{47}$$

In particular, by setting $t = 0$ in (47), we conclude from (11) that

$$\sum_x \pi(x)q_{xy} = 0, \qquad y \in \mathscr{S}. \tag{48}$$

It can be shown that (47) and (48) are valid even if $\mathscr{S}$ is an infinite set. Suppose, conversely, that (48) holds. If $\mathscr{S}$ is finite we conclude from the backward equation (14) that

$$\begin{aligned}
\frac{d}{dt}\sum_x \pi(x)P_{xy}(t) &= \sum_x \pi(x)P'_{xy}(t) \\
&= \sum_x \pi(x)\left(\sum_z q_{xz}P_{zy}(t)\right) \\
&= \sum_z \left(\sum_x \pi(x)q_{xz}\right) P_{zy}(t) \\
&= 0.
\end{aligned}$$

Thus

$$\sum_x \pi(x)P_{xy}(t)$$

is a constant in t and the constant value is given by

$$\sum_x \pi(x)P_{xy}(0) = \sum_x \pi(x)\delta_{xy} = \pi(y).$$

Consequently (46) holds. This conclusion is also valid if $\mathscr{S}$ is infinite, but the proof is much more complicated. In summary, (46) holds if and only if (48) holds.

A non-absorbing recurrent state x is called *positive recurrent* or *null recurrent* according as the *mean return time* $m_x = E_x(T_x)$ is finite or infinite. An absorbing state is considered to be positive recurrent. The process is said to be a *positive recurrent process* if all its states are positive recurrent and a *null recurrent process* if all its states are null recurrent. An irreducible recurrent process must be either a null recurrent process or a positive recurrent process. It can be shown that a stationary distribution is concentrated on the positive recurrent states, and hence a process that is transient or null recurrent has no stationary distribution. An irreducible positive recurrent process has a unique stationary distribution π, which, unless $\mathscr{S}$ consists of a single necessarily absorbing state, is given by

$$\pi(x) = \frac{1}{q_x m_x}, \qquad x \in \mathscr{S}. \tag{49}$$

Formula (49) is intuitively reasonable. For in a large time interval $[0, t]$, the process makes about t/m_x visits to x and the average time in x per visit is $1/q_x$. Thus the total time spent in state x during the time interval $[0, t]$ should be about $t/(q_x m_x)$ and the proportion of time spent in state x should be about $1/(q_x m_x)$. This argument can be made rigorous by using the strong law of large numbers as was done in Section 2.3.

Markov pure jump processes do not have any periodicities, and, in particular, for an irreducible positive recurrent process having stationary distribution π,

$$\lim_{t \to \infty} P_{xy}(t) = \pi(y), \qquad x, y \in \mathscr{S}. \tag{50}$$

If $X(0)$ has the initial distribution $\pi_0(x)$, $x \in \mathscr{S}$, then

$$P(X(t) = y) = \sum_x \pi_0(x) P_{xy}(t),$$

which, by (50) and the bounded convergence theorem, converges to

$$\sum_x \pi_0(x)\pi(y) = \pi(y)$$

as $t \to \infty$. In other words

$$\lim_{t \to \infty} P(X(t) = y) = \pi(y),$$

and hence the distribution of $X(t)$ converges to the stationary distribution π regardless of the initial distribution of the process.

3.3.1. Applications to birth and death processes.

Let $X(t)$, $0 \leq t < \infty$, be an irreducible birth and death process on $\{0, 1, 2, \ldots\}$. The process is transient if and only if the embedded birth and death chain having transition function $P(x, y) = Q_{xy}$, $x \geq 0$ and $y \geq 0$, is transient. From (18) in this chapter and the results in Section 1.7, we conclude that the birth and death process is transient if and only if

$$\sum_{x=1}^{\infty} \frac{\mu_1 \cdots \mu_x}{\lambda_1 \cdots \lambda_x} < \infty. \tag{51}$$

Equation (48) for a stationary distribution π becomes

$$\begin{aligned} &\pi(1)\mu_1 - \pi(0)\lambda_0 = 0, \\ &\pi(y + 1)\mu_{y+1} - \pi(y)\lambda_y = \pi(y)\mu_y - \pi(y - 1)\lambda_{y-1}, \qquad y \geq 1. \end{aligned} \tag{52}$$

It follows easily from (52) and induction that

$$\pi(y + 1)\mu_{y+1} - \pi(y)\lambda_y = 0, \qquad y \geq 0,$$

and hence that

$$\pi(y+1) = \frac{\lambda_y}{\mu_{y+1}} \pi(y), \qquad y \geq 0.$$

Consequently,

$$\pi(x) = \frac{\lambda_0 \cdots \lambda_{x-1}}{\mu_1 \cdots \mu_x} \pi(0), \qquad x \geq 1. \tag{53}$$

Set

$$\pi_x = \begin{cases} 1, & x = 0, \\ \dfrac{\lambda_0 \cdots \lambda_{x-1}}{\mu_1 \cdots \mu_x}, & x \geq 1. \end{cases} \tag{54}$$

Then (53) can be written as

$$\pi(x) = \pi_x \pi(0), \qquad x \geq 0. \tag{55}$$

Conversely, (52) follows from (54) and (55).

Suppose now that $\sum_x \pi_x < \infty$, i.e., that

$$\sum_{x=1}^{\infty} \frac{\lambda_0 \cdots \lambda_{x-1}}{\mu_1 \cdots \mu_x} < \infty. \tag{56}$$

We conclude from (55) that the birth and death process has a unique stationary distribution π, given by

$$\pi(x) = \frac{\pi_x}{\sum_{y=0}^{\infty} \pi_y}, \qquad x \geq 0. \tag{57}$$

If (56) fails to hold, the birth and death process has no stationary distribution.

In summary, an irreducible birth and death process on $\{0, 1, 2, \ldots\}$ is transient if and only if (51) holds, positive recurrent if and only if (56) holds, and null recurrent if and only if (51) and (56) each fail to hold, i.e., if and only if

$$\sum_{x=1}^{\infty} \frac{\mu_1 \cdots \mu_x}{\lambda_1 \cdots \lambda_x} = \infty \qquad \text{and} \qquad \sum_{x=1}^{\infty} \frac{\lambda_0 \cdots \lambda_{x-1}}{\mu_1 \cdots \mu_x} = \infty. \tag{58}$$

An irreducible birth and death process having finite state space $\{0, 1, \ldots, d\}$ is necessarily positive recurrent. It has a unique stationary distribution given by

$$\pi(x) = \frac{\pi_x}{\sum_{y=0}^{d} \pi_y}, \qquad 0 \leq x \leq d, \tag{59}$$

where π_x, $0 \leq x \leq d$, is given by (54).

Example 4. Show that the infinite server queue is positive recurrent and find its stationary distribution.

The infinite server queue has state space $\{0, 1, 2, \ldots\}$ and birth and death rates

$$\lambda_x = \lambda \qquad \text{and} \qquad \mu_x = x\mu, \qquad x \geq 0.$$

This process is clearly irreducible. It follows from (54) that

$$\pi_x = \frac{\lambda^x}{x!\,\mu^x} = \frac{(\lambda/\mu)^x}{x!}, \qquad x \geq 0.$$

Since

$$\sum_{x=0}^{\infty} \frac{(\lambda/\mu)^x}{x!} = e^{\lambda/\mu}$$

is finite, we conclude that the process is positive recurrent and has the unique stationary distribution π given by

$$\pi(x) = \frac{(\lambda/\mu)^x}{x!} e^{-\lambda/\mu}, \qquad x \geq 0, \tag{60}$$

which we note is a Poisson distribution with parameter λ/μ. We also note that (50) holds for this process, a direct consequence of (45) and (60).

Example 5. *N* server queue. Suppose customers arrive according to a Poisson process with parameter $\lambda > 0$. They are served by N servers, where N is a finite positive number. Suppose the service times are exponentially distributed with parameter μ and that whenever there are more than N customers waiting for service the excess customers form a queue and wait until their turn at one of the N servers. This process is a birth and death process on $\{0, 1, 2, \ldots\}$ with birth rates $\lambda_x = \lambda$, $x \geq 0$, and death rates

$$\mu_x = \begin{cases} x\mu, & 0 \leq x < N, \\ N\mu, & x \geq N. \end{cases}$$

Determine when this process is transient, null recurrent, and positive recurrent; and find the stationary distribution in the positive recurrent case.

Condition (51) for transience reduces to

$$\sum_{x=0}^{\infty} \left(\frac{N\mu}{\lambda}\right)^x < \infty.$$

Thus the N server queue is transient if and only if $N\mu < \lambda$. Condition (56) for positive recurrence reduces to

$$\sum_{x=0}^{\infty} \left(\frac{\lambda}{N\mu}\right)^x < \infty.$$

The N server queue is therefore positive recurrent if and only if $\lambda < N\mu$. Consequently the N server queue is null recurrent if and only if $\lambda = N\mu$. These results naturally are similar to those for the 1 server queue discussed in Chapters 1 and 2.

In the positive recurrent case,

$$\pi_x = \begin{cases} \dfrac{(\lambda/\mu)^x}{x!}, & 0 \le x < N, \\[2ex] \dfrac{(\lambda/\mu)^x}{N!\, N^{x-N}}, & x \ge N. \end{cases}$$

Set

$$K = \sum_{x=0}^{\infty} \pi_x = \sum_{x=0}^{N-1} \frac{(\lambda/\mu)^x}{x!} + \frac{(\lambda/\mu)^N}{N!}\left(1 - \frac{\lambda}{N\mu}\right)^{-1}.$$

We conclude that if $\lambda < N\mu$, the stationary distribution is given by

$$\pi(x) = \begin{cases} \dfrac{1}{K}\dfrac{(\lambda/\mu)^x}{x!}, & 0 \le x < N, \\[2ex] \dfrac{1}{K}\dfrac{(\lambda/\mu)^x}{N!\, N^{x-N}}, & x \ge N. \end{cases}$$

Exercises

1 Find the transition function of the two-state birth and death process by solving the forward equation.

2 Consider a birth and death process having three states 0, 1, and 2, and birth and death rates such that $\lambda_0 = \mu_2$. Use the forward equation to find $P_{0y}(t)$, $y = 0, 1, 2$.

Exercises 3–8 all refer to events occurring in time according to a Poisson process with parameter λ on $0 \le t < \infty$. Here $X(t)$ denotes the number of events that occur in the time interval $(0, t]$.

3 Find the conditional probability that there are m events in the first s units of time, given that there are n events in the first t units of time, where $0 \le m \le n$ and $0 \le s \le t$.

4 Let T_m denote the time to the mth event. Find the distribution function of T_m. *Hint:* $\{T_m \le t\} = \{X(t) \ge m\}$.

5 Find the density of the random variable T_m in Exercise 4. *Hint:* First consider some specific cases, say, $m = 1, 2, 3$.

6 Find $P(T_1 \le s \mid X(t) = n)$ for $0 \le s \le t$ and n a positive integer.

7 Let T be a random variable that is independent of the times when events occur. Suppose that T has an exponential density with parameter ν:

$$f_T(t) = \begin{cases} \nu e^{-\nu t}, & t > 0, \\ 0, & t \le 0. \end{cases}$$

Find the distribution of $X(T)$, which is the number of events occurring by time T. *Hint:* Use the formulas

$$P(X(T) = n) = \int_0^\infty f_T(t) P(X(T) = n \mid T = t)\, dt$$

and

$$P(X(T) = n \mid T = t) = P(X(t) = n).$$

8 Solve the previous exercise if T has the gamma density with parameters α and ν:

$$f_T(t) = \begin{cases} \nu^\alpha t^{\alpha-1} e^{-\nu t}/\Gamma(\alpha), & t > 0, \\ 0, & t \le 0. \end{cases}$$

9 Verify Equation (43).

10 Consider a pure death process on $\{0, 1, 2, \ldots\}$.
(a) Write the forward equation.
(b) Find $P_{xx}(t)$.
(c) Solve for $P_{xy}(t)$ in terms of $P_{x,y+1}(t)$.
(d) Find $P_{x,x-1}(t)$.
(e) Show that if $\mu_x = x\mu$, $x \ge 0$, for some constant μ, then

$$P_{xy}(t) = \binom{x}{y} (e^{-\mu t})^y (1 - e^{-\mu t})^{x-y}, \qquad 0 \le y \le x.$$

11 Let $X(t)$, $t \ge 0$, be the infinite server queue and suppose that initially there are x customers present. Compute the mean and variance of $X(t)$.

12 Consider a birth and death process $X(t)$, $t \ge 0$, such as the branching process, that has state space $\{0, 1, 2, \ldots\}$ and birth and death rates of the form

$$\lambda_x = x\lambda \qquad \text{and} \qquad \mu_x = x\mu, \qquad x \ge 0,$$

where λ and μ are nonnegative constants. Set

$$m_x(t) = E_x(X(t)) = \sum_{y=0}^{\infty} y P_{xy}(t).$$

(a) Write the forward equation for the process.
(b) Use the forward equation to show that $m_x'(t) = (\lambda - \mu) m_x(t)$.
(c) Conclude that

$$m_x(t) = x e^{(\lambda - \mu)t}.$$

13 Let $X(t)$, $t > 0$, be as in Exercise 12. Set

$$s_x(t) = E_x(X^2(t)) = \sum_{y=0}^{\infty} y^2 P_{xy}(t).$$

(a) Use the forward equation to show that

$$s_x'(t) = 2(\lambda - \mu)s_x(t) + (\lambda + \mu)m_x(t).$$

(b) Find $s_x(t)$.
(c) Find Var $X(t)$ under the condition that $X(0) = x$.

14 Suppose d particles are distributed into two boxes. A particle in box 0 remains in that box for a random length of time that is exponentially distributed with parameter λ before going to box 1. A particle in box 1 remains there for an amount of time that is exponentially distributed with parameter μ before going to box 0. The particles act independently of each other. Let $X(t)$ denote the number of particles in box 1 at time $t \geq 0$. Then $X(t)$, $t \geq 0$, is a birth and death process on $\{0, \ldots, d\}$.
(a) Find the birth and death rates.
(b) Find $P_{xd}(t)$. *Hint:* Let $X_i(t)$, $i = 0$ or 1, denote the number of particles in box 1 at time $t \geq 0$ that started in box i at time 0, so that $X(t) = X_0(t) + X_1(t)$. If $X(0) = x$, then $X_0(t)$ and $X_1(t)$ are independent and binomially distributed with parameters defined in terms of x and the transition function of the two-state birth and death process.
(c) Find $E_x(X(t))$.

15 Consider the infinite server queue discussed in Section 3.2.4. Let $X_1(t)$ and $X_2(t)$ be as defined there. Suppose that the initial distribution π_0 is a Poisson distribution with parameter ν.
(a) Use the formula

$$P(X_2(t) = k) = \sum_{x=k}^{\infty} \pi_0(x)P_x(X_2(t) = k)$$

to show that $X_2(t)$ has a Poisson distribution with parameter $\nu e^{-\mu t}$.
(b) Use the result of (a) to show that $X(t) = X_1(t) + X_2(t)$ has a Poisson distribution with parameter

$$\frac{\lambda}{\mu} + \left(\nu - \frac{\lambda}{\mu}\right) e^{-\mu t}.$$

(c) Conclude that $X(t)$ has the same distribution as $X(0)$ if and only if $\nu = \lambda/\mu$.

16 Consider a birth and death process on the nonnegative integers whose death rates are given by $\mu_x = x$, $x \geq 0$. Determine whether the process is transient, null recurrent, or positive recurrent if the birth rates are

(a) $\lambda_x = x + 1$, $x \geq 0$;
(b) $\lambda_x = x + 2$, $x \geq 0$.

17 Let $X(t)$, $t \geq 0$, be a birth and death process on the nonnegative integers such that $\lambda_x > 0$ and $\mu_x > 0$ for $x \geq 1$. Set $\gamma_0 = 1$ and

$$\gamma_x = \frac{\mu_1 \cdots \mu_x}{\lambda_1 \cdots \lambda_x}, \qquad x \geq 1.$$

(a) Show that if $\sum_{y=0}^{\infty} \gamma_y = \infty$, then $\rho_{x0} = 1$, $x \geq 1$.
(b) Show that if $\sum_{y=0}^{\infty} \gamma_y < \infty$, then

$$\rho_{x0} = \frac{\sum_{y=x}^{\infty} \gamma_y}{\sum_{y=0}^{\infty} \gamma_y}, \qquad x \geq 1.$$

Hint: Use Exercise 26 of Chapter 1.

18 Let $X(t)$, $t \geq 0$, be a single server queue ($N = 1$ in Example 5).
(a) Show that if $\mu \geq \lambda > 0$, then $\rho_{x0} = 1$, $x \geq 1$.
(b) Show that if $\mu < \lambda$, then

$$\rho_{x0} = (\mu/\lambda)^x, \qquad x \geq 1.$$

19 Consider the branching process introduced in Example 1. Use Exercise 17 to show that if $p \leq \frac{1}{2}$, then $\rho_{x0} = 1$ for all x and that if $p > \frac{1}{2}$, then

$$\rho_{x0} = \left(\frac{1-p}{p}\right)^x, \qquad x \geq 1.$$

20 Find the stationary distribution for the process in Exercise 14.

21 Suppose d machines are subject to failures and repairs. The failure times are exponentially distributed with parameter μ, and the repair times are exponentially distributed with parameter λ. Let $X(t)$ denote the number of machines that are in satisfactory order at time t. If there is only one repairman, then under appropriate reasonable assumptions, $X(t)$, $t \geq 0$, is a birth and death process on $\{0, 1, \ldots, d\}$ with birth rates $\lambda_x = \lambda$, $0 \leq x < d$, and death rates $\mu_x = x\mu$, $0 \leq x \leq d$. Find the stationary distribution for this process.

22 Consider a positive recurrent irreducible birth and death process on $\mathscr{S} = \{0, 1, 2, \ldots\}$, and let $X(0)$ have the stationary distribution π for its initial distribution. Then $X(t)$ has distribution π for all $t \geq 0$. The quantities

$$E\lambda_{X(t)} = \sum_{x=0}^{\infty} \lambda_x \pi(x) \qquad \text{and} \qquad E\mu_{X(t)} = \sum_{x=0}^{\infty} \mu_x \pi(x)$$

can be interpreted, respectively, as the average birth rate and the average death rate of the process.
(a) Show that the average birth rate equals the average death rate.
(b) What does (a) imply about a positive recurrent N server queue?

4 Second Order Processes

A *stochastic process* can be defined quite generally as any collection of random variables $X(t)$, $t \in T$, defined on a common probability space, where T is a subset of $(-\infty, \infty)$ and is usually thought of as the time parameter set. The process is called a *continuous parameter process* if T is an interval having positive length and a *discrete parameter process* if T is a subset of the integers. If $T = \{0, 1, 2, \ldots\}$ it is usual to denote the process by X_n, $n \geq 0$. The Markov chains discussed in Chapters 1 and 2 are discrete parameter processes, while the pure jump processes discussed in Chapter 3 are continuous parameter processes.

A stochastic process $X(t)$, $t \in T$, is called a *second order process* if $EX^2(t) < \infty$ for each $t \in T$. Second order processes and random variables defined in terms of them by various "linear" operations including integration and differentiation are the subjects of this and the next two chapters. We will obtain formulas for the means, variances, and covariances of such random variables.

We will consider continuous parameter processes almost exclusively in these three chapters. Since no new techniques are needed for handling the analogous results for discrete parameter processes, little would be gained by treating such processes in detail.

4.1. Mean and covariance functions

Let $X(t)$, $t \in T$, be a second order process. The *mean function* $\mu_X(t)$, $t \in T$, of the process is defined by

$$\mu_X(t) = EX(t).$$

The *covariance function* $r_X(s, t)$, $s \in T$ and $t \in T$, is defined by

$$r_X(s, t) = \operatorname{cov}(X(s), X(t)) = EX(s)X(t) - EX(s)EX(t).$$

This function is also called the auto-covariance function to distinguish it from the cross-covariance function which will be defined later. Since

$\operatorname{Var} X(t) = \operatorname{cov}(X(t), X(t))$, the variance of $X(t)$ can be expressed in terms of the covariance function as

$$(1) \qquad \operatorname{Var} X(t) = r_X(t, t), \qquad t \in T.$$

By a *finite linear combination* of the random variables $X(t)$, $t \in T$, we mean a random variable of the form

$$\sum_{j=1}^{n} b_j X(t_j),$$

where n is a positive integer, $t_1, \ldots, t_n$ are points in T, and $b_1, \ldots, b_n$ are real constants. The covariance between two such finite linear combinations is given by

$$\begin{aligned} \operatorname{cov}\left(\sum_{i=1}^{m} a_i X(s_i), \sum_{j=1}^{n} b_j X(t_j)\right) &= \sum_{i=1}^{m} \sum_{j=1}^{n} a_i b_j \operatorname{cov}(X(s_i), X(t_j)) \\ &= \sum_{i=1}^{m} \sum_{j=1}^{n} a_i b_j r_X(s_i, t_j). \end{aligned}$$

In particular,

$$(2) \qquad \operatorname{Var}\left(\sum_{j=1}^{n} b_j X(t_j)\right) = \sum_{i=1}^{n} \sum_{j=1}^{n} b_i b_j r_X(t_i, t_j).$$

It follows immediately from the definition of the covariance function that it is *symmetric* in s and t, i.e., that

$$(3) \qquad r_X(s, t) = r_X(t, s), \qquad s, t \in T.$$

It is also *nonnegative definite*. That is, if n is a positive integer, $t_1, \ldots, t_n$ are in T, and $b_1, \ldots, b_n$ are real numbers, then

$$\sum_{i=1}^{n} \sum_{j=1}^{n} b_i b_j r_X(t_i, t_j) \geq 0.$$

This is an immediate consequence of (2).

We say that $X(t)$, $-\infty < t < \infty$, is a *second order stationary process* if for every number τ the second order process $Y(t)$, $-\infty < t < \infty$, defined by

$$Y(t) = X(t + \tau), \qquad -\infty < t < \infty,$$

has the same mean and covariance functions as the $X(t)$ process. It is left as an exercise for the reader to show that this is the case if and only if $\mu_X(t)$ is independent of t and $r_X(s, t)$ depends only on the difference between s and t.

Let $X(t)$, $-\infty < t < \infty$, be a second order stationary process. Then

$$\mu_X(t) = \mu_X, \qquad -\infty < t < \infty,$$

where μ_X denotes the common mean of the random variables $X(t)$, $-\infty < t < \infty$. Since $r_X(s, t)$ depends only on the difference between s and t,

(4) $$r_X(s, t) = r_X(0, t - s), \qquad -\infty < s, t < \infty.$$

The function $r_X(t)$, $-\infty < t < \infty$, defined by

(5) $$r_X(t) = r_X(0, t), \qquad -\infty < t < \infty,$$

is also called the *covariance function* (or auto-covariance function) of the process. We see from (4) and (5) that

$$r_X(s, t) = r_X(t - s), \qquad -\infty < s, t < \infty.$$

It follows from (3) that $r_X(t)$ is symmetric about the origin, i.e., that

$$r_X(-t) = r_X(t), \qquad -\infty < t < \infty.$$

The random variables $X(t)$, $-\infty < t < \infty$, have a common variance given by

$$\operatorname{Var} X(t) = r_X(0), \qquad -\infty < t < \infty.$$

Recall Schwarz's inequality, which asserts that if X and Y are random variables having finite second moment, then $(EXY)^2 \le EX^2EY^2$. Applying Schwarz's inequality to the random variables $X - EX$ and $Y - EY$, we see that $(\operatorname{cov}(X, Y))^2 \le \operatorname{Var} X \operatorname{Var} Y$.

It follows from this last inequality that

$$|\operatorname{cov}(X(0), X(t))| \le \sqrt{\operatorname{Var} X(0) \operatorname{Var} X(t)},$$

and hence that

$$|r_X(t)| \le r_X(0), \qquad -\infty < t < \infty.$$

If $r_X(0) > 0$, the correlation between $X(s)$ and $X(s + t)$ is given independently of s by

$$\frac{\operatorname{cov}(X(s), X(s + t))}{\sqrt{\operatorname{Var} X(s)}\sqrt{\operatorname{Var} X(t)}} = \frac{r_X(t)}{r_X(0)}, \qquad -\infty < s, t < \infty.$$

Example 1. Let Z_1 and Z_2 be independent normally distributed random variables each having mean 0 and variance σ^2. Let λ be a real constant and set $X(t) = Z_1 \cos \lambda t + Z_2 \sin \lambda t$, $-\infty < t < \infty$. Find the mean and covariance functions of $X(t)$, $-\infty < t < \infty$, and show that it is a second order stationary process.

We observe first that

$$\mu_X(t) = EZ_1 \cos \lambda t + EZ_2 \sin \lambda t = 0, \qquad -\infty < t < \infty.$$

Next,

$$\begin{aligned} r_X(s, t) &= \operatorname{cov}(X(s), X(t)) \\ &= EX(s)X(t) - EX(s)EX(t) \\ &= EX(s)X(t) \\ &= E(Z_1 \cos \lambda s + Z_2 \sin \lambda s)(Z_1 \cos \lambda t + Z_2 \sin \lambda t) \\ &= EZ_1^2 \cos \lambda s \cos \lambda t + EZ_2^2 \sin \lambda s \sin \lambda t \\ &= \sigma^2(\cos \lambda s \cos \lambda t + \sin \lambda s \sin \lambda t) \\ &= \sigma^2 \cos \lambda(t - s). \end{aligned}$$

This shows that $X(t)$, $-\infty < t < \infty$, is a second order stationary process having mean zero and covariance function

$$r_X(t) = \sigma^2 \cos \lambda t, \qquad -\infty < t < \infty.$$

Example 2. Consider a two-state birth and death process as discussed in Section 3.2.1. It follows from that discussion that the transition probabilities of the process are given by

$$(6) \qquad \begin{aligned} P_{00}(t) &= 1 - P_{01}(t) = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} e^{-(\lambda+\mu)t}, \qquad t \geq 0, \\ P_{11}(t) &= 1 - P_{10}(t) = \frac{\lambda}{\lambda + \mu} + \frac{\mu}{\lambda + \mu} e^{-(\lambda+\mu)t}, \qquad t \geq 0, \end{aligned}$$

where λ and μ are positive constants. The process has the stationary distribution defined by

$$(7) \qquad \pi(0) = \frac{\mu}{\lambda + \mu} \quad \text{and} \quad \pi(1) = \frac{\lambda}{\lambda + \mu}.$$

In Chapter 3 we discussed birth and death processes defined on $0 \leq t < \infty$. Actually in the positive recurrent case it is possible to construct a corresponding process on $-\infty < t < \infty$ having the stationary distribution determined by (7). This process will be such that

$$(8) \qquad P(X(t) = 0) = \frac{\mu}{\lambda + \mu} \quad \text{and} \quad P(X(t) = 1) = \frac{\lambda}{\lambda + \mu}, \qquad -\infty < t < \infty,$$

and such that the Markov property

$$(9) \qquad P(X(t) = y \mid X(s) = x) = P_{xy}(t - s), \qquad -\infty < s \leq t < \infty,$$

holds, where $P_{xy}(t)$, $t \geq 0$, is given by (6). We will show that such a process is a second order stationary process and find its mean and covariance functions.

The mean function is given by

$$\begin{aligned}\mu_X(t) &= EX(t)\\ &= 0 \cdot P(X(t) = 0) + 1 \cdot P(X(t) = 1) = \frac{\lambda}{\lambda + \mu}.\end{aligned}$$

Let $-\infty < s \leq t < \infty$. Then

$$\begin{aligned}EX(s)X(t) &= P(X(s) = 1 \quad \text{and} \quad X(t) = 1)\\ &= P(X(s) = 1)P(X(t) = 1 \mid X(s) = 1)\\ &= P(X(s) = 1)P_{11}(t - s)\\ &= \frac{\lambda}{\lambda + \mu}\left(\frac{\lambda}{\lambda + \mu} + \frac{\mu}{\lambda + \mu} e^{-(\lambda+\mu)(t-s)}\right)\\ &= \left(\frac{\lambda}{\lambda + \mu}\right)^2 + \frac{\lambda\mu}{(\lambda + \mu)^2} e^{-(\lambda+\mu)(t-s)}.\end{aligned}$$

It follows that

$$r_X(s, t) = \frac{\lambda\mu}{(\lambda + \mu)^2} e^{-(\lambda+\mu)(t-s)}, \qquad -\infty < s \leq t < \infty.$$

By symmetry we see that

$$r_X(s, t) = \frac{\lambda\mu}{(\lambda + \mu)^2} e^{-(\lambda+\mu)|t-s|}, \qquad -\infty < s, t < \infty.$$

Thus $X(t)$, $-\infty < t < \infty$, is a second order stationary process having mean $\lambda/(\lambda + \mu)$ and covariance function

$$r_X(t) = \frac{\lambda\mu}{(\lambda + \mu)^2} e^{-(\lambda+\mu)|t|}, \qquad -\infty < t < \infty.$$

Other interesting examples of second order processes can be obtained from Poisson processes.

Example 3. Consider a Poisson process $X(t)$, $-\infty < t < \infty$, with parameter λ (see Section 3.2.2). This process satisfies the following properties:

(i) $X(0) = 0$.

(ii) $X(t) - X(s)$ has a Poisson distribution with mean $\lambda(t - s)$ for $s \leq t$.

(iii) $X(t_2) - X(t_1), X(t_3) - X(t_2), \ldots, X(t_n) - X(t_{n-1})$ are independent for $t_1 \leq t_2 \leq \cdots \leq t_n$.

We will now find the mean and covariance function of a process $X(t)$, $-\infty < t < \infty$, satisfying (i)–(iii). It follows from properties (i) and (ii) that $X(t)$ has a Poisson distribution with mean λt for $t \geq 0$ and $-X(t)$ has a Poisson distribution with mean $\lambda(-t)$ for $t < 0$. Thus

$$\mu_X(t) = \lambda t, \qquad -\infty < t < \infty.$$

Since the variance of a Poisson distribution equals its mean, we see that $X(t)$ has finite second moment and that $\operatorname{Var} X(t) = \lambda|t|$. Let $0 \leq s \leq t$. Then

$$\operatorname{cov}(X(s), X(s)) = \operatorname{Var} X(s) = \lambda s.$$

It follows from properties (i) and (iii) that $X(s)$ and $X(t) - X(s)$ are independent, and hence

$$\operatorname{cov}(X(s), X(t) - X(s)) = 0.$$

Thus

$$\begin{aligned}\operatorname{cov}(X(s), X(t)) &= \operatorname{cov}(X(s), X(s) + X(t) - X(s)) \\ &= \operatorname{cov}(X(s), X(s)) + \operatorname{cov}(X(s), X(t) - X(s)) \\ &= \lambda s.\end{aligned}$$

If $s < 0$ and $t > 0$, then by properties (i) and (iii) the random variables $X(s)$ and $X(t)$ are independent, and hence

$$\operatorname{cov}(X(s), X(t)) = 0.$$

The other cases can be handled similarly. We find in general that

$$r_X(s, t) = \begin{cases} \lambda \min(|s|, |t|), & st \geq 0, \\ 0, & st < 0. \end{cases} \tag{10}$$

The process from Example 3 is not a second order stationary process. In the next example we will consider a closely related process which is a second order stationary process.

Example 4. Let $X(t)$, $-\infty < t < \infty$, be a Poisson process with parameter λ. Set

$$Y(t) = X(t + 1) - X(t), \qquad -\infty < t < \infty.$$

Find the mean and covariance function of the $Y(t)$ process, and show that it is a second order stationary process.

Since $EX(t) = \lambda t$, it follows that

$$\begin{aligned}EY(t) &= E(X(t + 1) - X(t)) \\ &= \lambda(t + 1) - \lambda t = \lambda,\end{aligned}$$

so the random variables $Y(t)$ have common mean λ. To compute the covariance function of the $Y(t)$ process, we observe that if $|t - s| \geq 1$, then the random variables $X(s + 1) - X(s)$ and $X(t + 1) - X(t)$ are independent by property (iii). Consequently,

$$r_Y(s, t) = 0 \qquad \text{for} \qquad |t - s| \geq 1.$$

Suppose $s \leq t < s + 1$. Then

$$\begin{aligned} \operatorname{cov}(Y(s), Y(t)) &= \operatorname{cov}(X(s + 1) - X(s), X(t + 1) - X(t)) \\ &= \operatorname{cov}(X(t) - X(s) + X(s + 1) - X(t), X(s + 1) \\ &\qquad - X(t) + X(t + 1) - X(s + 1)). \end{aligned}$$

It follows from property (iii) and the assumptions on s and t that

$$\operatorname{cov}(X(t) - X(s), X(s + 1) - X(t)) = 0,$$

$$\operatorname{cov}(X(t) - X(s), X(t + 1) - X(s + 1)) = 0,$$

and

$$\operatorname{cov}(X(s + 1) - X(t), X(t + 1) - X(s + 1)) = 0.$$

By property (ii)

$$\begin{aligned} \operatorname{cov}(X(s + 1) - X(t), X(s + 1) - X(t)) &= \operatorname{Var}(X(s + 1) - X(t)) \\ &= \lambda(s + 1 - t). \end{aligned}$$

Thus

$$\operatorname{cov}(Y(s), Y(t)) = \lambda(s + 1 - t).$$

By using symmetry we find in general that

$$r_Y(s, t) = \begin{cases} \lambda(1 - |t - s|), & |t - s| < 1, \\ 0, & |t - s| \geq 1. \end{cases}$$

Thus $Y(t)$, $-\infty < t < \infty$, is a second order stationary process having mean λ and covariance function

$$r_Y(t) = \begin{cases} \lambda(1 - |t|), & |t| < 1, \\ 0, & |t| \geq 1. \end{cases}$$

In Figure 1 we have graphed the covariance function for three different second order stationary processes. These covariance functions are special cases of those found in Examples 1, 2, and 4 respectively. In each case $r_X(0) = 1$ and hence $r_X(t)$ is equal to the correlation between $X(0)$ and $X(t)$. In the top curve of Figure 1 we see that the correlation oscillates between -1 and 1. In the middle curve the correlation decreases exponentially fast as $|t| \to \infty$. In the bottom curve the correlation decreases linearly to zero as $|t|$ increases from 0 to 1 and remains zero for all larger values of $|t|$.

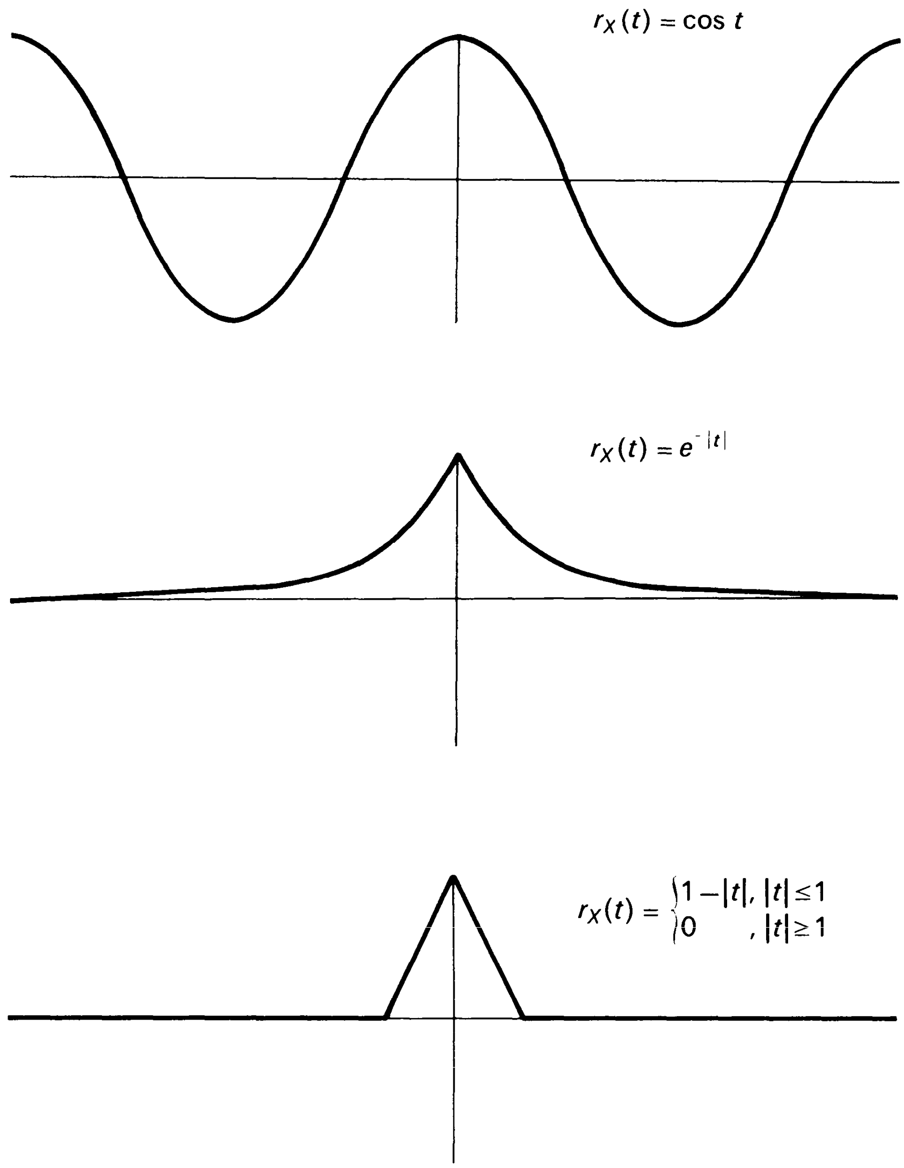


Figure 1

Consider two second order processes $X(t)$, $t \in T$, and $Y(t)$, $t \in T$. Their *cross-covariance function* is defined as

$$r_{XY}(s, t) = \operatorname{cov}(X(s), Y(t)), \qquad s, t \in T.$$

Clearly

$$r_{XY}(s, t) = r_{YX}(t, s)$$

and

$$r_{XX}(s, t) = r_X(s, t).$$

The cross-covariance function can be used to find the covariance function of the sum of two processes. Indeed,

$$\begin{aligned} r_{X+Y}(s, t) &= \operatorname{cov}(X(s) + Y(s), X(t) + Y(t)) \\ &= r_{XX}(s, t) + r_{XY}(s, t) + r_{YX}(s, t) + r_{YY}(s, t), \end{aligned}$$

which can be rewritten as

$$(11) \qquad r_{X+Y}(s, t) = r_X(s, t) + r_{XY}(s, t) + r_{YX}(s, t) + r_Y(s, t).$$

In the important case when the cross-covariance function vanishes, (11) reduces to

$$r_{X+Y}(s, t) = r_X(s, t) + r_Y(s, t). \tag{12}$$

These formulas are readily extended to sums of any finite number of processes. Consider in particular n second order stationary processes $X_1(t)$, $-\infty < t < \infty, \ldots, X_n(t)$, $-\infty < t < \infty$, whose cross-covariance functions all vanish. Then their sum

$$X(t) = X_1(t) + \cdots + X_n(t), \qquad -\infty < t < \infty,$$

is a second order stationary process such that

$$\mu_X = \sum_{k=1}^{n} \mu_{X_k} \tag{13}$$

and

$$r_X(t) = \sum_{k=1}^{n} r_{X_k}(t), \qquad -\infty < t < \infty. \tag{14}$$

Example 5. Let $Z_{11}, Z_{12}, Z_{21}, Z_{22}, \ldots, Z_{n1}, Z_{n2}$ be $2n$ independent normally distributed random variables each having mean zero and such that

$$\operatorname{Var} Z_{k1} = \operatorname{Var} Z_{k2} = \sigma_k^2, \qquad k = 1, \ldots, n.$$

Let $\lambda_1, \ldots, \lambda_n$ be real constants and set

$$X(t) = \sum_{k=1}^{n} (Z_{k1} \cos \lambda_k t + Z_{k2} \sin \lambda_k t), \qquad -\infty < t < \infty.$$

Find the mean and covariance functions of $X(t)$, $-\infty < t < \infty$.

Set

$$X_k(t) = Z_{k1} \cos \lambda_k t + Z_{k2} \sin \lambda_k t.$$

It follows from the independence of the Z's that the cross-covariance function between $X_i(t)$ and $X_j(t)$ vanishes for $i \neq j$. Thus by using (13) and (14) together with the results of Example 1, we see that $X(t)$, $-\infty < t < \infty$, is a second order stationary process having mean zero and covariance function

$$r_X(t) = \sum_{k=1}^{n} \sigma_k^2 \cos \lambda_k t, \qquad -\infty < t < \infty. \tag{15}$$

4.2. Gaussian processes

A stochastic process $X(t)$, $t \in T$, is called a *Gaussian process* if every finite linear combination of the random variables $X(t)$, $t \in T$, is normally distributed. (In this context constant random variables are regarded as normally distributed with zero variance.) Gaussian processes are also called normal processes, and normally distributed random variables are sometimes said to have a Gaussian distribution. If $X(t)$, $t \in T$, is a Gaussian process, then for each $t \in T$, $X(t)$ is normally distributed and, in particular, $EX^2(t) < \infty$. Thus a Gaussian process is necessarily a second order process. Gaussian processes have many nice theoretical properties that do not hold for second order processes in general. They are also widely used in applications, especially in engineering and in the physical sciences.

Example 6. Show that the process $X(t)$, $-\infty < t < \infty$, from Example 1 is a Gaussian process.

To verify that this is a Gaussian process, we let n be a positive integer and choose real numbers $t_1, \ldots, t_n$ and $a_1, \ldots, a_n$. Now

$$X(t) = Z_1 \cos \lambda t + Z_2 \sin \lambda t,$$

where Z_1 and Z_2 are independent and normally distributed. Thus

$$\begin{aligned} &a_1 X(t_1) + \cdots + a_n X(t_n) \\ &\quad = Z_1(a_1 \cos \lambda t_1 + \cdots + a_n \cos \lambda t_n) + Z_2(a_1 \sin \lambda t_1 + \cdots + a_n \sin \lambda t_n) \end{aligned}$$

is a linear combination of independent normally distributed random variables and therefore is itself normally distributed.

It is left as an exercise for the reader to show that the process in Example 5 is also a Gaussian process.

Two stochastic processes $X(t)$, $t \in T$, and $Y(t)$, $t \in T$, are said to have the same joint distribution functions if for every positive integer n and every choice of $t_1, \ldots, t_n$, all in T, the random variables

$$X(t_1), \ldots, X(t_n)$$

have the same joint distribution function as the random variables

$$Y(t_1), \ldots, Y(t_n).$$

One of the most useful properties of Gaussian processes is that if two such processes have the same mean and covariance functions, then they also have the same joint distribution functions. We omit the proof of this result. To see that the Gaussian assumption is necessary, observe that the process defined in Exercise 15 has the same mean and covariance functions as that from Example 1 with $\sigma^2 = 1$ but not the same joint distribution functions.

The mean and covariance functions can also be used to find the higher moments of a Gaussian process.

Example 7. Let $X(t)$, $t \in T$, be a Gaussian process having zero means. Find $EX^4(t)$ in terms of the covariance function of the process.

We recall that if X is normally distributed with mean 0 and variance σ^2, then $EX^4 = 3\sigma^4$. Since $X(t)$ is normally distributed with mean 0 and variance $r_X(t, t)$, we see that

$$EX^4(t) = 3(r_X(t, t))^2.$$

Let n be a positive integer and let $X_1, \ldots, X_n$ be random variables. They are said to have a joint *normal* (or *Gaussian*) *distribution* if

$$a_1X_1 + \cdots + a_nX_n$$

is normally distributed for every choice of the constants $a_1, \ldots, a_n$. A stochastic process $X(t)$, $t \in T$, is a Gaussian process if and only if for every positive integer n and every choice of $t_1, \ldots, t_n$ all in T, the random variables $X(t_1), \ldots, X(t_n)$ have a joint normal distribution.

Let $X_1, \ldots, X_n$ be random variables having a joint normal distribution and a density f with respect to integration on R^n. (Such a density exists if and only if the covariance matrix of $X_1, \ldots, X_n$ has nonzero determinant.) It can be shown that f is necessarily of the form

$$f(x_1, \ldots, x_n) = \frac{1}{(2\pi)^{n/2}(\det \Sigma)^{1/2}} \exp\left[-\tfrac{1}{2}(\mathbf{x} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right], \tag{16}$$

where Σ is the covariance matrix

$$\Sigma = \begin{bmatrix} \operatorname{cov}(X_1, X_1) \cdots \operatorname{cov}(X_1, X_n) \\ \vdots \qquad\qquad \vdots \\ \operatorname{cov}(X_n, X_1) \cdots \operatorname{cov}(X_n, X_n) \end{bmatrix},$$

$\mathbf{x}$ and $\boldsymbol{\mu}$ are the vectors

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \qquad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix},$$

and $'$ denotes matrix transpose. In particular, if $n = 2$, then (16) can be written as

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - \rho^2}} \exp\left[-\frac{Q(x_1, x_2)}{2}\right], \tag{17}$$

where

$$Q(x_1, x_2) = \frac{1}{(1-\rho^2)} \times \left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right].$$

Here μ_1 and σ_1^2 denote the mean and variance of X_1, μ_2 and σ_2^2 denote the mean and variance of X_2, and ρ denotes the correlation between X_1 and X_2. One can also use (16) to show that the conditional expectation of X_n given $X_1, \ldots, X_{n-1}$ is a linear function of these $n-1$ random variables, i.e., that

$$E[X_n \mid X_1 = x_1, \ldots, X_{n-1} = x_{n-1}] = a + b_1 x_1 + \cdots + b_{n-1}x_{n-1}$$

for suitable constants $a, b_1, \ldots, b_{n-1}$.

A stochastic process $X(t)$, $-\infty < t < \infty$, is said to be *strictly stationary* if for every number τ the stochastic process $Y(t)$, $-\infty < t < \infty$, defined by

$$Y(t) = X(t+\tau), \qquad -\infty < t < \infty,$$

has the same joint distribution functions as the $X(t)$ process. A strictly stationary process need not have finite second moments and hence need not be a second order process. It is clear, however, that if a strictly stationary process does have finite second moments, then it is a second order stationary process. The converse is not true in general. It is left as an exercise for the reader to demonstrate by an example that a second order stationary process need not be strictly stationary.

Let $X(t)$, $-\infty < t < \infty$, be a second order stationary process which is also a Gaussian process. Then this process is necessarily strictly stationary. For if τ is any real number, then the $Y(t)$ process defined by $Y(t) = X(t+\tau)$, $-\infty < t < \infty$, is a Gaussian process having the same mean and covariance functions as the $X(t)$ process. It therefore has the same joint distribution functions as the $X(t)$ process.

Since the processes in Examples 1 and 5 are Gaussian and second order stationary, they are also strictly stationary. The second order stationary processes from Examples 2 and 4 are not Gaussian, but it can be shown that they too are strictly stationary.

4.3. The Wiener process

It has long been known from microscopic observations that particles suspended in a liquid are in a state of constant highly irregular motion. It gradually came to be realized that the cause of this motion is the bombardment of the particles by the smaller invisible molecules of the liquid. Such motion is called "Brownian motion," named after one of the first scientists to study it carefully.

Many mathematical models for this physical process have been proposed. We will now describe one such model. Let the location of a particle be described by a Cartesian coordinate system whose origin is the location of the particle at time $t = 0$. Then the three coordinates of the position of the particle vary independently, each according to a stochastic process $W(t)$, $-\infty < t < \infty$, satisfying the following properties:

(i) $W(0) = 0$.
(ii) $W(t) - W(s)$ has a normal distribution with mean 0 and variance $\sigma^2(t - s)$ for $s \leq t$.
(iii) $W(t_2) - W(t_1), W(t_3) - W(t_2), \ldots, W(t_n) - W(t_{n-1})$ are independent for $t_1 \leq t_2 \leq \cdots \leq t_n$.

Here σ^2 is some positive constant.

Property (i) follows from our choice of the coordinate system. Properties (ii) and (iii) are plausible if the motion is caused by an extremely large number of unrelated and individually negligible collisions which have no more tendency to move the particle in one direction than in the opposite direction. In particular, the central limit theorem makes it reasonable to suppose that the increments $W(t) - W(s)$ are normally distributed.

This model was initiated, in a different form, by Albert Einstein in 1905. He related the parameter σ^2 to various physical parameters including Avogadro's number. Estimation of σ^2 together with other measurements in a scientific experiment conducted shortly thereafter led to an estimate of Avogadro's number that is within 19 percent of the presently accepted value. Einstein's work and its experimental confirmation gave added evidence for the atomic basis of matter, which was still being questioned at the turn of the century.

Although the mathematical model is reasonable and fits the experimental data quite well, it has certain theoretical deficiencies that will be discussed in Section 5.3. In Chapter 6 we will discuss another mathematical model for the physical process.

A stochastic process $W(t)$, $-\infty < t < \infty$, satisfying properties (i)–(iii) is called the *Wiener process* with parameter σ^2. Mathematicians Norbert Wiener and Paul Lévy developed much of the theory, and the process is also known as the Wiener-Lévy process and as Brownian motion. The Wiener process is usually assumed to satisfy an additional property involving "continuity of the sample functions," which we will discuss in Section 5.1.2.

It follows immediately from the properties of the Wiener process that the random variables $W(t)$ all have mean 0 and that

$$E(W(t_2) - W(t_1))(W(t_4) - W(t_3)) = 0, \quad t_1 \le t_2 \le t_3 \le t_4. \tag{18}$$

The covariance function of the process is

$$r_W(s, t) = \begin{cases} \sigma^2 \min(|s|, |t|), & st > 0, \\ 0, & st \le 0. \end{cases} \tag{19}$$

The proof of (19) is virtually identical to that of Formula (10) for the covariance function of the Poisson process defined in Example 3. It is left as an exercise for the reader to show that

$$\begin{aligned} E(W(s) - W(a))(W(t) - W(a)) \\ = \sigma^2 \min(s - a, t - a), \quad s \ge a \text{ and } t \ge a. \end{aligned} \tag{20}$$

The Wiener process is a Gaussian process. In other words, if $t_1 \le \cdots \le t_n$ and $b_1, \ldots, b_n$ are real constants, the random variable

$$b_1 W(t_1) + \cdots + b_n W(t_n)$$

is normally distributed. In proving this result we can assume, with no loss of generality, that one of the numbers $t_1, \ldots, t_n$, say t_k, equals zero. Then each of the random variables $W(t_1), \ldots, W(t_n)$ is a linear combination of the increments $W(t_2) - W(t_1), \ldots, W(t_n) - W(t_{n-1})$. Indeed, $W(t_k) = 0$,

$$W(t_j) = (W(t_{k+1}) - W(t_k)) + \cdots + (W(t_j) - W(t_{j-1})), \qquad k < j \le n,$$

and

$$W(t_j) = (W(t_j) - W(t_{j+1})) + \cdots + (W(t_{k-1}) - W(t_k)), \qquad 1 \le j < k.$$

Thus $b_1 W(t_1) + \cdots + b_n W(t_n)$ can also be written as a linear combination of the increments $W(t_2) - W(t_1), \ldots, W(t_n) - W(t_{n-1})$. Now these increments are independent and normally distributed. Thus any linear combination of them, in particular,

$$b_1 W(t_1) + \cdots + b_n W(t_n)$$

is normally distributed.

Exercises

1 Let $X(t)$, $-\infty < t < \infty$, be a second order process. Show that it is a second order stationary process if and only if $\mu_X(t)$ is independent of t and $r_X(s, t)$ depends only on the difference between s and t.

2 Let $X(t)$, $-\infty < t < \infty$, be a second order process. Show that it is a second order stationary process if and only if $EX(s)$ and $EX(s)X(s + t)$ are both independent of s.

3 Let $X(t)$, $-\infty < t < \infty$, be a second order stationary process and set $Y(t) = X(t + 1) - X(t)$, $-\infty < t < \infty$. Show that the $Y(t)$ process is a second order stationary process having zero means and covariance function

$$r_Y(t) = 2r_X(t) - r_X(t - 1) - r_X(t + 1).$$

4 Let $X(t)$, $-\infty < t < \infty$, be a second order stationary process.
(a) Show that

$$\operatorname{Var}(X(s + t) - X(s)) = 2(r_X(0) - r_X(t)).$$

(b) Show that for $M > 0$

$$P(|X(s + t) - X(s)| \geq M) \leq \frac{2}{M^2}(r_X(0) - r_X(t)).$$

5 Let $X(t)$, $-\infty < t < \infty$, be a Poisson process with parameter λ and set $Y(t) = X(t) - tX(1)$, $0 \leq t \leq 1$. Find the mean and covariance functions of the $Y(t)$ process.

6 Let $U_1, \ldots, U_n$ be independent random variables, each uniformly distributed on $(0, 1)$. Let $\psi(t, x)$, $0 \leq t \leq 1$ and $0 \leq x \leq 1$, be defined by

$$\psi(t, x) = \begin{cases} 1, & x \leq t, \\ 0, & x > t. \end{cases}$$

Then

$$X(t) = \frac{1}{n} \sum_{k=1}^{n} \psi(t, U_k), \qquad 0 \leq t \leq 1,$$

is the *empirical distribution function* of $U_1, \ldots, U_n$. Compute the mean and covariance functions of the $X(t)$ process.

7 Let $X(t)$, $-\infty < t < \infty$, be a second order stationary process having covariance function $r_X(t)$, $-\infty < t < \infty$. Set $Y(t) = X(t + 1)$, $-\infty < t < \infty$. Find the cross-covariance function between the $X(t)$ process and the $Y(t)$ process.

8 Let R and Θ be independent random variables such that Θ is uniformly distributed on $[0, 2\pi)$ and R has the density

$$f_R(r) = \begin{cases} \dfrac{r}{\sigma^2} e^{-r^2/2\sigma^2}, & 0 < r < \infty, \\ 0, & r \leq 0, \end{cases}$$

where σ is a positive constant. It follows by using the change of variable formula involving Jacobians that $R \cos \Theta$ and $R \sin \Theta$ are independent random variables, each normally distributed with mean 0 and variance σ^2. Let λ be a positive constant and set

$$X(t) = R \cos(\lambda t + \Theta), \qquad -\infty < t < \infty.$$

Show that the $X(t)$ process is a second order stationary process having mean zero and covariance function

$$r_X(t) = \sigma^2 \cos \lambda t, \qquad -\infty < t < \infty.$$

9 Let $R_1, \ldots, R_n, \Theta_1, \ldots, \Theta_n$ be independent random variables such that the Θ's are uniformly distributed on $[0, 2\pi)$ and R_k has the density

$$f_{R_k}(r) = \begin{cases} \dfrac{r}{\sigma_k^2} e^{-r^2/2\sigma_k^2}, & 0 < r < \infty, \\ 0, & r \le 0, \end{cases}$$

where $\sigma_1, \ldots, \sigma_n$ are positive constants. Let $\lambda_1, \ldots, \lambda_n$ be positive constants and set

$$X(t) = \sum_{k=1}^{n} R_k \cos(\lambda_k t + \Theta_k).$$

Show that the $X(t)$ process is a second order stationary process having mean zero and covariance function

$$r_X(t) = \sum_{k=1}^{n} \sigma_k^2 \cos \lambda_k t.$$

10 Show that the $X(t)$ process in Example 5 is a Gaussian process.

11 Show that the $X(t)$ process in Exercise 9 is a Gaussian process.

12 Let $X(t)$, $-\infty < t < \infty$, be a Gaussian process and let f and g be functions from $(-\infty, \infty)$ to $(-\infty, \infty)$. Show that $Y(t) = f(t)X(g(t))$, $-\infty < t < \infty$, is a Gaussian process and find its mean and covariance functions.

13 Let $X(t)$, $-\infty < t < \infty$, be a Gaussian process having mean zero and set $Y(t) = X^2(t)$, $-\infty < t < \infty$.
(a) Find the mean and covariance functions of the $Y(t)$ process.
(b) Show that if the $X(t)$ process is a second order stationary process, then so is the $Y(t)$ process.

14 Let X_1 and X_2 have the joint density given by (17).
(a) Find the conditional density of X_2 given $X_1 = x_1$.
(b) Find the conditional expectation of X_2 given $X_1 = x_1$.

15 Let Z_1 and Z_2 be independent and identically distributed random variables taking on the values -1 and 1 each with probability 1/2. Show that $X(t) = Z_1 \cos \lambda t + Z_2 \sin \lambda t$, $-\infty < t < \infty$, is a second order stationary process which is not strictly stationary.

In the remaining problems $W(t)$, $-\infty < t < \infty$, is the Wiener process with parameter σ^2.

16 Verify Formula (20).

17 Find the distribution of $W(1) + \cdots + W(n)$ for a positive integer n. *Hint:* Use the formulas

$$1 + 2 + \cdots + n = \frac{n(n+1)}{2}$$

and

$$1^2 + 2^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}.$$

18 Set

$$X(t) = \frac{W(t+\varepsilon) - W(t)}{\varepsilon}, \qquad -\infty < t < \infty,$$

where ε is a positive constant. Show that the $X(t)$ process is a stationary Gaussian process having covariance function

$$r_X(t) = \begin{cases} \dfrac{\sigma^2}{\varepsilon}\left(1 - \dfrac{|t|}{\varepsilon}\right), & |t| < \varepsilon, \\ 0, & |t| \geq \varepsilon. \end{cases}$$

19 Set

$$X(t) = e^{-\alpha t} W(e^{2\alpha t}), \qquad -\infty < t < \infty,$$

where α is a positive constant. Show that the $X(t)$ process is a stationary Gaussian process having covariance function

$$r_X(t) = \sigma^2 e^{-\alpha|t|}, \qquad -\infty < t < \infty.$$

20 Find the mean and covariance functions of the following processes:

(a) $X(t) = (W(t))^2, \quad t \geq 0;$

(b) $X(t) = tW(1/t), \quad t > 0;$

(c) $X(t) = c^{-1}W(c^2 t), \quad t \geq 0;$

(d) $X(t) = W(t) - tW(1), \quad 0 \leq t \leq 1.$